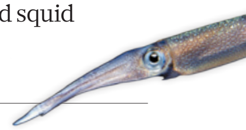
THIS WEEK

EDITORIALS



An accident waiting to happen

The release of radioactive material at a US nuclear-waste repository reveals an all-too-common picture of complacency over safety and a gradual downgrading of regulations.

On St Valentine's Day, the United States' flagship geological repository for nuclear waste dodged a bullet. Deep below the New Mexico desert, something went wrong. One or more drums of nuclear waste ruptured, probably because of a chemical reaction or explosion. Thousands of drums are held in the 655-metre deep underground repository, designed to safely contain for thousands of years the low- and medium-level radioactive remnants of US military programmes. Just 15 years after it opened, the Waste Isolation Pilot Plant (WIPP) near Carlsbad has been hurriedly closed down while officials seek answers.

Parts of the repository were contaminated with long-lived transuranic radioactive elements, including americium and plutonium. The extent of the contamination is still being established, but the amounts released were not small, and last week officials announced that the repository will remain closed for at least 18 months and possibly much longer. A small amount of radioactivity was also vented to the surface, and 21 workers were exposed to what seem to have been low levels.

It is clear that both the accident and its consequences could have been much worse. Maintenance resulting from a separate and unrelated accident on 5 February — a vehicle fire underground — meant that from 6 to 10 February the ventilation was unfiltered, and real-time continuous radiation monitors were switched off. Had the accident happened then, rather than on 14 February, the release would only have been detected during manual radiation readings that are taken each morning, meaning that workers would unknowingly have been exposed, and higher levels of radioactivity would have reached the environment.

On the evening of the accident, a continuous radiation monitor underground, which sounded the alert to high radiation levels in a waste-storage area at 11.14 p.m., was the only one in service, as all the others were out of order. This resulted in automatic switching of the ventilation to pass by high-efficiency particulate air (HEPA) filtration to catch radioactive particles. Shortly after the alert, a vigilant shift manager opened large fans to vent the repository contamination through the HEPA filters to the environment; this should have happened automatically with no need for manual intervention — but it had been switched to manual some years ago. The ventilation system also fell short of nuclear-safety norms, as it had gaps that allowed some radiation to reach the environment. Workers plugged these gaps with high-density foam on 6 March.

The mantra for WIPP was to "start clean and stay clean". Accidents, the government said, would never happen. But as a News article on page 267 details, a Department of Energy (DOE) report on the incident outlines how fanciful that promise was. The report describes an atmosphere of complacency. It lists a litany of failings, from an insidious continual deregulation of safety standards and cutting of corners, to dilapidated safety equipment, and a lax security culture. WIPP's response to the accident itself was "delayed and ineffective" adds the report.

The consequences of a release of radioactivity at WIPP, a repository for low- and medium-level waste deep underground in a remote region, are much less serious than those at a nuclear power plant. But as with the Fukushima nuclear power plant in Japan, the same characteristic errors were in play: hubris, overconfidence in safety assumptions, dilution or non-respect of safety standards, a weak security culture and, crucially, lack of tough, independent scientific and technical oversight.

And, as at Fukushima, it took an accident to uncover glaring safety weaknesses and the lack of a strong safety culture — an essential element in safe nuclear operation. The DOE, which operates WIPP, and the WIPP regulators —including the Environmental Protection Agency — seem to have been asleep at the wheel. The uncovering of these safety deficiencies is all the more disconcerting given that the authorities have been proposing to expand WIPP from a site for low- and medium-level waste to one that would also hold both high-level surplus weapons-grade plutonium and much hotter spent nuclear fuel.

"It took an accident to uncover glaring safety weaknesses and the lack of a strong safety culture."

In the past, WIPP was a model of how to integrate science into the planning and design of a nuclear-waste repository, and how to gain public confidence in that science. Its recent shortcomings are a further blow to the pressing need to find ways to deal safely with the vast quantities of accumulated defence and civilian wastes. WIPP and planned repositories elsewhere would do well to heed the lessons of WIPP's troubles, and strive to ensure that transparent independent scientific oversight of projects is made a top priority and maintained. ■

Full support

Germany should follow the United Kingdom's lead and spell out the benefits of animal research.

Scientists in the United Kingdom have reason to be grateful this week, after research institutions came together to pledge greater public support for researchers who use animals in their work. The UK 'concordat' sets out how institutions that undertake animal research will publicize it. Signatories, which include major charities, drug companies and universities, say that they will increase the amount of information they provide about what happens in their laboratories to inform the public about the value of animal research, and will report annually on how they are moving to greater openness.

It is a laudable aim, and scientists in another European country must be wondering what they need to do to earn similar support. While the

United Kingdom was putting the final touches to its concordat, six newspapers in Germany were running a full-page advertisement questioning whether scientists who experiment on animals are even human.

The advert opens with the quote: "Animal experimenters are a particular type of creature — one should not casually call them human." It publishes a photograph of primate researcher Andreas Kreiter of the University of Bremen, a long-standing target of campaigners in the country, and describes him as a tormenter of animals whose research is without value. The advert closes with calls for citizens to treat all animal experimenters with contempt and denounce their work publicly.

Last week, the powerful Alliance of Science Organisations in Germany declared in a press statement that the lobby group that placed the adverts, Tierversuchsgegner Bundesrepublik Deutschland, had crossed acceptable boundaries. The alliance's strong words represent a welcome change from its unhelpful default policy of keeping its head below the parapet. But German scientists deserve more.

Now that it has broken its long silence over the use of animals in research, the alliance cannot retreat. It should follow the UK example and push for wider public awareness. Given the political weight of the institutions it represents — the Max Planck Society, the Leopoldina national academy, the universities and the Helmholtz Association among them — such a stance could make a crucial difference.

Scientists across Germany have been lobbying for nearly three years for the alliance to create a web resource for journalists and the public that makes available the true facts about research using animals. The Max Planck Society, which is taking the lead in a dragged-out effort to gather data about the value of such a resource, has doubts. But this should proceed as soon as possible.

The Tierversuchsgegner's advertising campaign may have been expressly designed to provoke a response, to keep the subject of animal research in the media. That is all the more reason for the alliance to collate an accessible pool of information for the public.

An immediate goal could be to prevent a recurrence of the advert, which ran in publications including the quality intellectual nationals *Die Zeit* and the *Frankfurter Allgemeine Zeitung*. What were they thinking?

Germany takes the right of freedom of expression very seriously. But newspapers must balance this right with the first clause of Germany's 1949 constitution, which states that the dignity of humans is inviolable. This was designed to ensure that a regime could never again label people 'subhuman', and so unworthy of life, as the Nazis did.

This is not the first time that such disturbing terminology has been levelled at science in Germany. At a public lecture in March, the award-winning novelist Sibylle Lewitscharoff attacked reproductive medicine, and referred to people born by artificial insemination as 'half-creatures'.

"To make their points, animal-rights groups often lie or omit key information."

The use of such aggressive language in debates about the ethical limits to medical research is worrying. When it comes to the use of animals in science, it underlines the importance of a proactive public stance. The most fiery animal-rights groups may be small, but they amplify their messages by appealing to people's emotions. To make their points, they often lie or omit key information about the tight regulation and oversight of animal experiments. Journalists have no ready source of counter-information. Research agencies have been nervous of commenting openly, fearing that it might open more scientists to attack. Many medical charities avoid mentioning that they support research with animals for fear of putting off donors.

In 2010, frustrated academic and industry researchers created the Basel Declaration, whose signatories commit to speaking publicly about their work and the value of experiments with animals. More than 2,300 individuals around the world have signed up — 431 of them in Germany — and 13 institutes and societies have given their support. Still, it remains a relatively small effort, and relies on donations to cover its costs. The UK concordat represents a more powerful tool that other countries, Germany chief among them, should emulate. ■

Hard data

It has been no small feat for the Protein Data Bank to stay relevant for 100,000 structures.

Sherlock Holmes understood: "It is a capital mistake," he said, "to theorise before one has data." Data are the lifeblood of science, the foundation of innovation. Behind every great discovery is a pile of data; but, crucially, it should not be too far behind.

For more than four decades, the Protein Data Bank (PDB) has been where structural biologists keep their data close. Nearly every biology-publishing journal in the world, *Nature* included, requires protein structures to be deposited in the PDB before publication.

So there was considerable worry at the database when *Nature* accepted a molecular map of HIV's capsid protein shell last year (G. Zhao *et al. Nature* **497,** 643–646; 2013). The multimillion-atom complex was larger than anything then in the PDB, and the database's team had to devise a way to make the data dump available (and useful) at short notice.

Thus it goes at the PDB — whose trove surpasses 100,000 structures this week (see page 265) — and other long-running archives that have managed to stay relevant and essential. It is not easy. Just ask the scientists, funders, technicians and others who shepherd them.

Money is often the limiting factor. Computer storage and processing power may be getting cheap as chips, but much of the expense is in paying the people (many of them highly trained scientists) who organize and verify data entries, and engage scientific communities.

There are many ways for a database to stay in the black. The three-decades-old GenBank, a clearing house for DNA sequences, is funded directly by the US government's support of the National Center for Biotechnology Information (NCBI). By contrast, the 50-year-old Cambridge Structural Database, which stores 700,000 small-molecule structures, gets by on support from industry and around 1,300 institutes.

The PDB is actually hosted by several organizations that provide access to the same data trove, each funded independently. Gerard Kleywegt, who heads the European franchise at the European Bioinformatics Institute (EBI) in Hinxton, UK, says that healthy competition between his portal and others in the United States and Japan helps him to get grants, and keeps the database pertinent. Scientists "vote with their mouse clicks", he says. "They go to the place where they get the best answer for their questions."

In the 1970s, protein structures were consumed by a small community of X-ray crystallographers interested in the nitty-gritty of individual enzymes. Now scientists use a range of techniques to determine structures, and researchers of many stripes want to know how proteins behave in a larger context, such as in a malignant cancer cell. A database must change with the times, or face extinction.

The closure of a database is not so awful — as long as its useful information remains available elsewhere. In 2011, NCBI announced that it was mothballing a database that collected information about protein fragments used in proteomics experiments. A competing database run by the EBI has since swallowed up those data. But with 100,147 structures (as *Nature* went to press), and growing at about 200 per week, the PDB, at least, shows no sign of folding. ■

➲ NATURE.COM
To comment online, click on Editorials at: go.nature.com/xhunqv

WORLD VIEW *A personal take on events*

VANESSA MINTEER

Is it right to reverse extinction?

Several groups are working to bring back long-dead species, but these efforts could undo some hard-learned lessons, argues **Ben Minteer**.

The North American sky, according to historical accounts, was once black with passenger pigeons. Hunters, however, saw to it that the sky was clear of the birds by the second half of the nineteenth century. 'Martha', the last individual of the species, expired in the Cincinnati Zoo in 1914.

Writers have long elegized this vanished bird. The great conservationist-philosopher Aldo Leopold issued the most poignant tribute in his 1949 book *A Sand County Almanac*: "We grieve," he wrote, "because no living man will see again the onrushing phalanx of victorious birds, sweeping a path for spring across the March skies, chasing the defeated winter from all the woods and prairies of Wisconsin."

But what if we could once again see those victorious birds sweeping their path across the March skies?

Leopold could not have known that only a handful of decades after he wrote these words we would be on the verge of a scientific revolution in efforts to reverse the death of species. The 'de-extinction' movement — a prominent group of scientists, futurists and their allies — argues that we no longer have to accept the finality of extinction. By applying techniques such as cloning and genetic engineering, they believe that we can and should return lost species such as the passenger pigeon to the landscape. This is the goal of the San Francisco, California-based Long Now Foundation, which is actively supporting scientific efforts to recreate the lost bird within its 'Revive & Restore' project. But it does not stop there. Scientists in Spain say they are close to cloning the Pyrenean ibex, a mountain goat that took its last breath in 2000. Other species have also been targeted, including the Tasmanian tiger and even the woolly mammoth.

THE 'DE-EXTINCTION' MOVEMENT ARGUES THAT WE NO LONGER HAVE TO ACCEPT THE FINALITY OF EXTINCTION.

The de-extinction lobby makes persuasive arguments. The most powerful among them appeal to our sense of justice: de-extinction is our opportunity to right past wrongs and to atone for our moral failings. Advocates also point to the sense of wonder that the revival of extinct species could encourage among the public. Although we will always have passenger pigeons in museums and books, "book-pigeons," Leopold lamented, "cannot dive out of a cloud to make the deer run for cover, nor clap their wings in thunderous applause of mast-laden woods." De-extinctionists argue further that the revived species will restore lost ecological functions and enhance the diversity of ecosystems.

At the same time, the de-extinction proposal raises considerable concerns. Resuscitated species could create problems in contemporary environments and for native species that have evolved in the absence of the vanished biota. As with the introduction of any species into a new environment, there are risks of disease transmission and biological invasion. Some conservationists also express the fear that, given decades of ecological change and human development, the landscape won't be able to support the revived populations.

➲ NATURE.COM
Discuss this article online at:
go.nature.com/tqazw8

Others fret about the limited genetic diversity of any 'de-extinguished' species and question the assumption that reviving a genome is the same thing as recovering the animal's behaviour and identity, which evolved over millennia. And there is also the particularly distressing concern that such aggressive manipulation of wildlife might actually end up diminishing our desire (and our limited resources) to conserve extant species — and that it would entail harmful interference in the lives of animals.

The most troubling aspect of de-extinction, however, is what it might mean for us. Attempting to revive lost species is in many ways a refusal to accept our moral and technological limits in nature. De-extinction thus reflects a new kind of Promethean spirit that attempts to leverage our boundless cleverness and powerful tools for conservation rather than for human enhancement. But things did not end very well for Prometheus.

Leopold was aware of our tendency to let our gadgets get out in front of our ethics. "Our tools," he cautioned in the late 1930s, "are better than we are, and grow better faster than we do. They suffice to crack the atom, to command the tides. But they do not suffice for the oldest task in human history: to live on a piece of land without spoiling it." The real challenge is to live more lightly on the land and to address the moral and cultural forces that drive unsustainable and ecologically destructive practices.

That is why there is great virtue in keeping extinct species extinct. Meditation on their loss reminds us of our fallibility and our finitude. We are a wickedly smart species, and occasionally a heroic and even exceptional one. But we are a species that often becomes mesmerized by its own power.

It would be silly to deny the reality of that power. But we should also cherish and protect the capacity of nature, including those parts of nature that are no longer with us, to teach us something profound about the value of collective self-restraint and human limits. Few things teach us this sort of earthly modesty any more.

It cuts against the progressive aims of science to say it, but there can be wisdom in taking our foot off the gas, in resisting the impulse to further control and manipulate; to fix nature.

Ben A. Minteer *is a professor of environmental ethics at Arizona State University in Tempe. A longer version of this essay will appear in his forthcoming book (co-edited with Steve Pyne),* After Preservation: Saving American Nature in the Age of Humans.
e-mail: ben.minteer@asu.edu

RESEARCH HIGHLIGHTS

Selections from the scientific literature

PHYSICS

Tractor beam pulls in objects

An array of ultrasound beams can drag centimetre-sized objects towards it.

Mike McDonald at the University of Dundee, UK, Gabriel Spalding at Illinois Wesleyan University in Bloomington and their colleagues sculpted interference patterns in the array so that much of the acoustic energy bounced off the sides or rear of an object in front of the array. This drove the object towards the ultrasound sources. The effect has been previously shown with light waves, but sound waves can move larger objects.

Such control might prove useful in non-invasive surgery: for example, it could be used to manipulate drug-delivery packages inside the body or to precisely cut out tumours.

Phys. Rev. Lett. 112, **174302 (2014)**

ROGER T. HANLON

ANIMAL BIOLOGY

Injury shapes squid behaviour

Squid that are sensitized to pain by injury are quicker to flee from predators, showing an adaptive benefit to injury and pain.

Robyn Crook and Edgar Walters of the University of Texas Medical School at Houston and their colleagues took several squid (*Doryteuthis pealeii*; pictured) and inflicted a minor injury on one arm of each animal. When exposed to black sea bass, the previously injured squid fled or hid from these predators earlier than uninjured animals. But squid that were treated with anaesthetics before the injury, and so did not develop neural sensitization, failed to change their behaviour. As a result, these animals were less likely to survive encounters with the predator than injured individuals that were not anaesthetized. This is the first experimental evidence that pain-like neural sensitization is an adaptive response to injury, the authors say.

Curr. Biol. **http://doi.org/sp8 (2014)**

GEOPHYSICS

How El Niño slows the planet's spin

The El Niño Pacific weather event affects how long the day is, but two types of El Niño do this in two different ways.

Weather changes affect the planet's rotation speed, and thus day length, by changing the atmosphere's pressure over topographical features. A team led by Olivier de Viron, now at the University of La Rochelle in France, studied atmospheric behaviour between 1948 and 2013.

The researchers found that when El Niños make Pacific waters warmer in the east, they set up strong pressure gradients above big mountain ranges (such as the Andes) that increase the time it takes the planet to spin by slightly more than 0.1 millisecond. By contrast, El Niños with warmer central Pacific waters produce only about half as much Earth-changing drag.

Geophys. Res. Lett. **http://doi.org/snq (2014)**

ECOLOGY

Longlines better for deep seas

Fishing with longlines has little effect on the vulnerable ecosystems of the deep sea, according to Telmo Morato and his team at the University of the Azores in Horta, Portugal.

Deep-sea fishing practices such as trawling have proved controversial owing to concerns about damage to slow-growing species at the bottom of the ocean. The researchers studied data from longline fishing, a technique that uses one main line with many shorter, hooked lines attached, around the Azores islands, and compared them to published data on the effects of bottom trawling. They estimate that between 4,000 and 23,000 longline deployments would be needed to remove 90% of cold-water corals in a given area, compared with just 13 trawls.

Regulated longline fishing could be a more sustainable method of deep-sea fishing than trawling, the authors suggest.

Sci. Rep. 4, **4837 (2014)**

ATMOSPHERIC SCIENCE

Detecting rainfall from the bottom up

A method that allows researchers to estimate global rainfall levels using soil-moisture data could help to improve hazard planning for floods and landslides.

To estimate rainfall in places that lack ground-based rain gauges, researchers rely on satellite data of atmospheric moisture, but this is notoriously inaccurate. Luca Brocca at the National Research Council in Perugia, Italy, and his colleagues developed an algorithm that calculates rainfall amounts on the basis of satellite data on soil moisture. They compared their estimates with rain-gauge data and found that their method accurately estimates rainfall in several regions around the world.

Moreover, their algorithm is better than a state-of-the-art method at detecting light rainfall events and precipitation at high latitudes.

***J. Geophys. Res. Atmos.* http://doi.org/sp7 (2014)**

NAWAZISH NAQVI

MATERIALS

Graphene analogue carries current

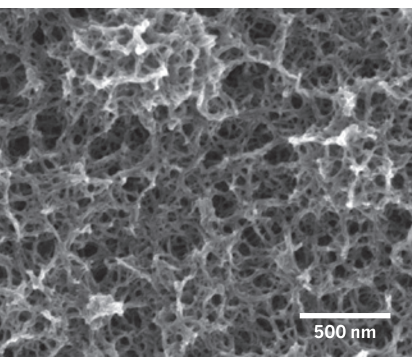
A self-assembling polymer that forms thin films and conducts electricity could beat graphene as a candidate material for flexible electronics.

Graphene, made of an atom-thick sheet of carbon, is flexible but cannot be used as a semiconductor in transistors because it lacks a 'band gap'. Mircea Dincă at the Massachusetts Institute of Technology in Cambridge and his colleagues mixed nickel with an organic compound called HITP and ammonia in water to produce a graphene-like structure with the important band gap.

The ingredients self-assemble into a flat, honeycomb-like structure (**pictured**) that has excellent electrical conductivity, unlike most other self-assembled organic–inorganic systems. The team studied the material only in bulk form, but say that the results could be even better if the polymer was in two-dimensional sheets, perhaps leading to more efficient solar cells and supercapacitors.

***J. Am. Chem. Soc.* http://doi.org/spj (2014)**

DENNIS SHEBERLA



GENOMICS

When brown and polar bears split

Polar bears evolved adaptations specific to the Arctic in fewer than 20,500 generations, and diverged from brown bears much more recently than is sometimes claimed.

Rasmus Nielsen at the University of California, Berkeley, and his colleagues sequenced the genomes of 79 polar bears (*Ursus maritimus*) and 10 brown bears (*Ursus arctos*) and found that the two species diverged between 343,000 and 479,000 years ago.

Many of the genes under the greatest selection pressure in the polar bear are associated with the cardiovascular system. In particular, this bear seems to have evolved modifications in its vascular system that allow the animal to tolerate an extremely fatty diet made up mostly of blubbery seal meat.

***Cell* http://doi.org/sp3 (2014)**

For a longer story on this research, see go.nature.com/zovyry

SOCIAL SELECTION

Maths reality check resonates online

Biologists of all stripes are sharing an essay by Harvard University mathematician-turned-biologist Jeremy Gunawardena that makes a sobering observation: the mathematical equations at the core of many biological models fail to reflect nature. He argues that the components of all quantitative models should be verifiable and, most of all, the conclusions should be falsifiable. Or, in his words: "Stick the model's neck out." Jason Moore, a geneticist at Dartmouth College in New Hampshire, tweeted: "This paper is so good I am actually printing it out"— high praise in the paperless age.

***BMC Biol.* 12, 29 (2014)**

Based on data from altmetric.com. Altmetric is supported by Macmillan Science and Education, which owns Nature Publishing Group.

NATURE.COM
For more on popular papers:
go.nature.com/mpqjve

ASTROPHYSICS

Big planets could alter star rotation

Massive planets with close-in orbits — also known as hot Jupiters — may influence the rotation and surface activity of their host stars.

Katja Poppenhaeger and Scott Wolk at the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts, analysed the emissions of binary-star systems, in which only one of the two stars in the system hosted an exoplanet. Comparing the differences between the emissions of the stars in each pair allowed the authors to measure the influence of the exoplanet on its host star. Using X-ray data from the Chandra and XMM-Newton space telescopes, the researchers found that the stars hosting hot Jupiters showed more magnetic activity than their planet-free companions.

Magnetic activity increases with rotation, so the authors suggest that the gravitational influence of the hot Jupiters may have counteracted the natural slowing of their host stars' spin over time.

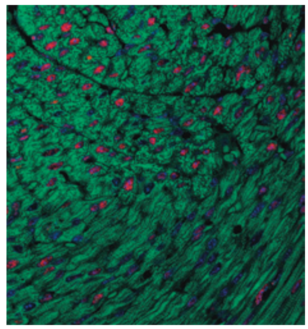
***Astron. Astrophys.* 565, L1 (2014)**

BIOLOGY

Thyroid makes young hearts grow

A surge of thyroid hormone just before adolescence causes mouse hearts to grow drastically, suggesting that the organ may be easier to regenerate than previously thought.

Ahsan Husain of Emory University School of Medicine in Atlanta, Georgia, and Robert Graham of the Victor Chang Cardiac Research Institute in Sydney, Australia, and their colleagues labelled heart muscle cells of baby mice with a chemical. When the mice were 15 days old, the number of cardiomyocytes (**pictured**, red) increased by about 40%.



It had previously been thought that cardiomyocytes stopped replicating just after birth. The findings suggest that giving thyroid hormone to babies with heart defects might help to repair the organ.

***Cell* 157, 795–807 (2014)**

NATURE.COM
For the latest research published by *Nature* visit:
www.nature.com/latestresearch

SEVEN DAYS *The news in brief*

POLICY

Baby-blood law

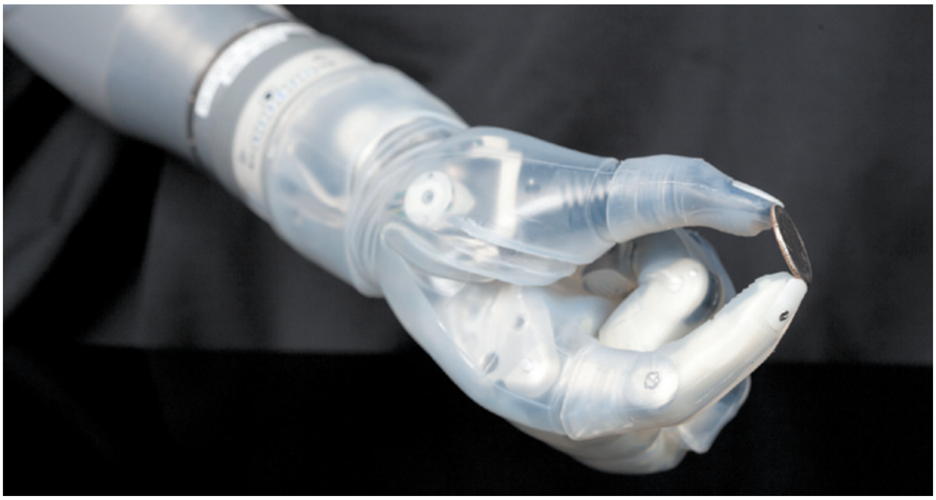
Minnesota will once again allow blood spots collected from newborns to be kept and used for research, unless parents opt out. On 6 May, state governor Mark Dayton signed the controversial bill into law, reversing a 2011 ruling by the state's Supreme Court, which said that the practice violated state laws that require written, informed consent for the collection and storage of genetic information. The 2011 ruling allowed most blood spots to be stored for only 71 days to allow time for routine disease screening, and the state was forced to destroy more than 1 million samples. See go.nature.com/5ckmbm for more.

Stanford axes coal

Stanford University in Palo Alto, California, will no longer invest in coal-mining companies from its US$18.7-billion endowment fund, it announced on 6 May. The move follows the recommendation of the university's advisory panel of students, staff and alumni, which reviewed the social and environmental implications of investing in fossil fuels. Stanford is the largest of a number of US universities that have elected to remove fossil-fuel stock from their investments. The university said that the value of its investments in coal mining was "small".

GM labels

On 8 May, Vermont became the first US state to mandate labelling of food containing genetically modified ingredients by July 2016. Representatives of the food and biotechnology industries condemned the law, and the US Grocery Manufacturers Association in Washington DC pledged to challenge it in the federal court. Vermont's attorney general said that he is prepared to launch a vigorous defence. More than 60 countries require labelling of genetically modified foods.



DARPA

FDA approves high-tech prosthetic arm

US regulators have approved the first prosthetic arm that can perform complex movements by picking up on electrical signals sent to muscles by the brain. The US Food and Drug Administration (FDA) in Silver Spring, Maryland, gave the DEKA Arm System (pictured) the green light on 9 May. The device uses electrodes to detect electrical activity caused by muscle contraction near the prosthesis. The arm enables some amputees to perform more-complex activities, such as using keys and locks and preparing food, than are possible with current prosthetic technology. The system was a test case for a fast-track FDA programme announced in 2011 to speed up approvals for medical technologies.

RESEARCH

Misconduct verdict

Haruko Obokata, a stem-cell scientist at Japan's RIKEN Center for Developmental Biology in Kobe, who was charged with research misconduct, has lost her appeal to have her case reviewed. The RIKEN institute confirmed on 8 May that it has advised Obokata to retract two papers she published in *Nature* describing a new method to reprogram cells to an embryonic state. On 7 May, an investigation committee advised RIKEN to deny Obokata's appeal. See go.nature.com/rttlvk for more.

Animal wrongs

An alliance of ten leading German research organizations has spoken out against animal-rights activists who are targeting neuroscientist Andreas Kreiter at the University of Bremen. Kreiter conducts research on monkeys. On 7 May, the alliance said it "expressly and decisively condemns" an advertising campaign that personally attacks Kreiter and that suggests he, and other animal experimenters, should not be thought of as human. See page 259 and go.nature.com/lzphx5 for more.

Open doors

Scientists who work with animals in the United Kingdom have pledged to be less guarded about their activities. On 14 May, 72 organizations including universities, charities, drug companies and government funders released a 'concordat' committing to more openness than in the past. On 1 May, the UK government proposed jettisoning a rule that has prevented it from releasing much of the information it holds on animal research (see go.nature.com/zijvk2).

US climate changes

Climate change is already affecting the United States, warns a 6 May report from the nation's government. The country's third national

assessment of climate change impacts says that rising greenhouse-gas emissions have made US summers longer and winters shorter, and have upped the risk of extreme weather events. John Holdren, President Barack Obama's chief science adviser, called for "urgent action to combat the threats to Americans from climate change". See go.nature.com/frmuuz for more.

Biosafety law

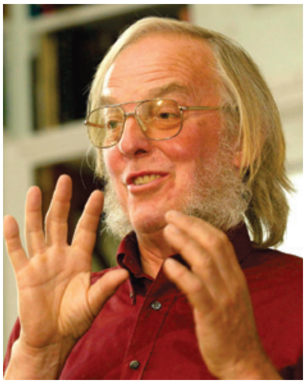
The German government should bring in a new law to regulate potentially dangerous bioscience research, advised a national ethics council in a report released on 7 May. Such dual-use research, which includes studies on lethal pathogens and toxins, should be subject to approval by a federal, interdisciplinary commission of experts, the council said. The report also recommended that German universities and research organizations should set up a national code of conduct. The German government commissioned the report two years ago.

PEOPLE

Beagle 2 leader dies

British planetary scientist Colin Pillinger (**pictured**), best known for his role as lead scientist on the Beagle 2 mission to Mars, died on

PHOTOSHOT



7 May, aged 70. Beagle 2 lost touch with Earth after reaching the red planet on Christmas Day 2003, but the mission propelled Pillinger into the limelight and he quickly became an ambassador for UK space science. He started his career studying Apollo-mission lunar samples at NASA, then worked at the University of Cambridge, UK, before moving to Britain's Open University in Milton Keynes, where he spent 35 years.

NOAA chief

Oceanographer Richard Spinrad is the new chief scientist at the US National Oceanic and Atmospheric Administration (NOAA). Spinrad, appointed by President Barack Obama on 8 May, is the first person to hold the job since 1996. Congress blocked Obama's first attempt to revive the chief scientist slot in 2010. Spinrad is no stranger to NOAA: from 2005 to 2010 he served as an associate administrator there overseeing oceanic and atmospheric research.

FACILITIES

European lasers

A third facility in eastern Europe's Extreme Light Infrastructure (ELI), a network that will allow scientists worldwide to probe the frontiers of laser science, received a funding green light on 8 May. The European Commission approved €111 million (US$153 million) from the European Regional Development Fund so that Hungary can build the ELI Attosecond Light Pulse Source near the University of Szeged. The fund, designed to help poor regions to improve their infrastructures, has already paid for the ELI's two other pillars — the ELI Nuclear Physics facility near Bucharest and the ELI Beamlines facility near Prague.

Lost at sea

An US$8-million deep-sea-research craft belonging to the Woods Hole Oceanographic Institution in Massachusetts has been wrecked at sea. The unmanned vehicle Nereus was lost 9,990 metres under water while exploring the Kermadec Trench off New Zealand on 10 May. Crew members from the ship *Thomas G. Thompson*, who were operating Nereus, later recovered debris from the sea surface. The submersible seems to have imploded, the institute said in a statement. See go.nature.com/qiwrmd for more.

BUSINESS

Pharma exchange

US pharmaceutical company Merck announced on 6 May that it is selling its consumer care business, which includes over-the-counter pharmaceuticals, to Germany's Bayer for US$14.2 billion. Merck will also pay Bayer at least $1 billion to share in the development, marketing and profits of a class of cardiovascular drugs called soluble guanylate cyclase inhibitors. Bayer is developing the drugs to treat heart failure and some forms of pulmonary hypertension.

COMING UP

13–16 MAY
The United Nations holds its first meeting to address the problem of lethal autonomous weapons systems, or 'killer robots', in Geneva, Switzerland. The meeting will include a debate between leading robotics experts.
go.nature.com/lugvpj

22–25 MAY
Elections for the European parliament take place. Science issues that may play a part include support for stem-cell research and genetically modified crops.
go.nature.com/lahzaj

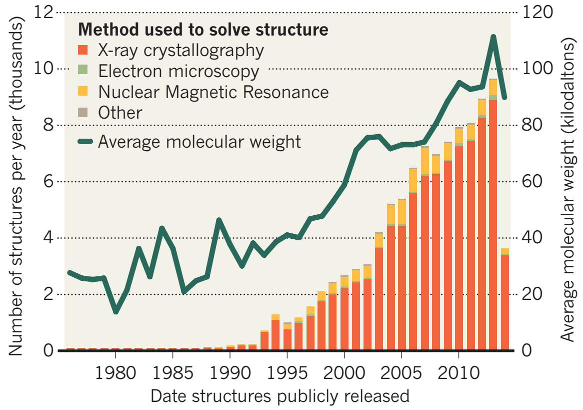
TREND WATCH

A digital compendium of proteins and other biomolecules has surpassed 100,000 entries, with the release of 219 new structures on 14 May. The Protein Data Bank (PDB) was started in 1971 to store three-dimensional structural data down to the atomic level. Then and now, scientists mapped most proteins using X-ray crystallography, but they are increasingly using other tools, such as nuclear magnetic resonance and electron microscopy. See also page 260.

SOURCE: PDBE

ONE HUNDRED THOUSAND PROTEIN STRUCTURES

Biomolecular structures stored in the Protein Data Bank are getting bigger and more complex.



➲ NATURE.COM
For daily news updates see:
www.nature.com/news

WIPP

Stored packages at the Waste Isolation Pilot Plant were inspected for signs of damage after an accident in February.

NUCLEAR WASTE

Call for better oversight of nuclear-waste storage

Accident at US repository highlights need for tougher safety monitoring, say experts.

BY DECLAN BUTLER

A serious accident in February at the United States' only deep-storage repository for nuclear waste might never have happened had the government not disbanded a key independent scientific body charged with oversight of the safety of the facility.

The Waste Isolation Pilot Plant (WIPP), carved out of a salt bed 655 metres below the desert near Carlsbad in New Mexico, is run by the Department of Energy (DOE) and stores low- and medium-level military nuclear waste, containing long-lived, man-made elements such as plutonium and americium. But there are politically controversial plans to store far hotter high-level waste at the site. Nuclear-waste experts say that the accident — in which a container is thought to have ruptured or exploded — along with management errors and a lack of oversight at WIPP, highlight the need for an independent risk assessment of any proposed expansion.

The facility was opened in 1999 and is designed to operate for a few decades, after which it will be sealed forever. The accident on 14 February released moderate levels of radioactivity into the repository, as well as small amounts into the environment, and officials say that the plant will not reopen for at least 18 months.

According to a preliminary report released on 24 April by a DOE-appointed Accident Investigation Board, the root cause of the accident lies with the department's field office and Nuclear Waste Partnership, the contractor ▶

that operates the site, both in Carlsbad. They failed to identify radiological risks and make plans to control them, the report's authors said. They added that maintenance of safety systems was neglected, and that DOE oversight was "ineffective".

The report's findings are in sharp contrast to WIPP's past record as a model of how to safely design and operate a deep geological waste repository. Many scientists attribute that reputation to the tough oversight provided until 2004 by the Environmental Evaluation Group (EEG), a scientific body that was set up in 1978 and charged with protecting public health and the environment.

The EEG was staunchly independent of the DOE, and its technical expertise and authority were widely viewed as key to the public and political trust that the repository won. The Blue Ribbon Commission on America's Nuclear Future, a government scientific advisory group, said in its 2012 report that the EEG "provided an independent and credible source" of information and review of WIPP.

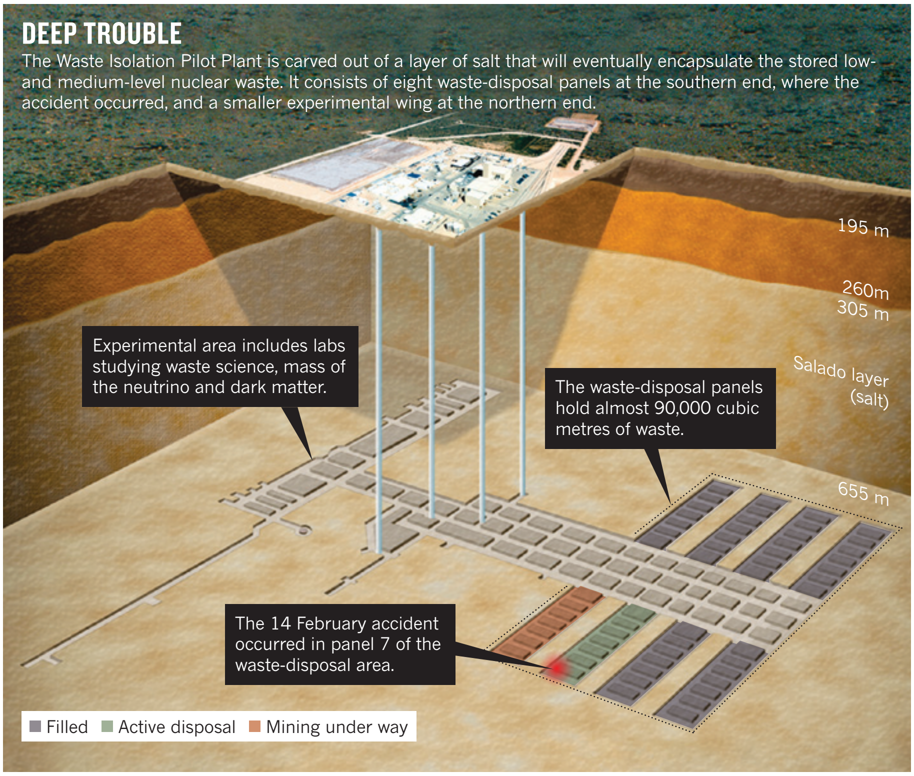
But in 2004, with WIPP by then fully operational, the group was defunded and disbanded. Responsibility for oversight moved primarily to the New Mexico Environment Department in Santa Fe and the US Environmental Protection Agency. "With the demise of EEG, independent technical and scientific oversight and transparency of WIPP has diminished," says George Anastas, a radiation and nuclear-safety consultant, and a former EEG staff scientist.

The accident report put into focus some of the decisions taken since the changeover.

In 2006, for example, WIPP watered down a requirement that all waste containers have their contents analysed to characterize the type of waste they hold and to verify that they do not contain flammable, corrosive or reactive materials. This might have had a direct bearing on the accident. Although the ultimate cause has yet to be determined, an inspection on 30 April ruled out a roof or wall collapse, and experts say that photographs showing evidence of heat damage in panel 7, where the accident occurred (see 'Deep trouble'), are consistent with an explosion of one or more containers.

And in 2009, WIPP eliminated 15 of the 22 potential accidents it had previously been required to withstand, "without any clear justification", the report said.

Under the EEG's watch, the reduction in postulated accidents would not have happened, says Lokesh Chaturvedi, an engineering geologist who was deputy director of the EEG from 1982 to 2000. "I have no doubt in my mind that had EEG continued, the standard for inspecting the drums before shipping to WIPP would not have been diluted," he adds. James Channell, an environmental engineer and health physicist who worked at the EEG for 21 years, says that the accident and the report highlight the need to immediately reinstate an oversight body akin to the EEG. That body's first job should be to carry out an independent review of the accident, he says.



Channell adds that proposals to expand WIPP's remit from storing just defence-related low- and medium-level waste to include high-level waste also should not be approved without rigorous scientific evaluation.

One such proposal has been floated by the US National Nuclear Security Administration for the disposal of 34 tonnes of high-level, weapons-grade plutonium. A report released on 29 April by the agency concluded that storage at WIPP was the cheapest of several options, costing \$8.8 billion. Other plans aim to store high-level spent fuel from nuclear power plants, which is much hotter than the high-level military waste. The proposals stem from another repository problem: the government had originally planned to store high-level waste in a facility at Yucca Mountain in Nevada, but the project was mothballed in 2010.

"Independent technical and scientific oversight and transparency of WIPP has diminished."

Research and scientific consensus on the safety of storing very hot high-level waste in salt beds are lacking. In particular, the effects of high temperatures on the salt are not well characterized. Heat might draw water out of salt crystals towards the waste, for example, potentially creating danger from steam and pressure, says Don Hancock, director of nuclear-waste safety at the Southwest Research and Information Center, a watchdog group in Albuquerque, New Mexico.

The DOE and other scientists argue, however, that such salt, heat and water interactions may be inconsequential, and that the heat might actually have the positive effect of driving moisture out of repositories. Heat also speeds up salt creep, which could help to encapsulate waste faster, but might also create operational problems. Only field experiments, some of which have started at WIPP, will be able to definitively demonstrate that salt is a safe medium for storing high-level waste that generates large amounts of heat, they say.

Ed Lyman, a nuclear expert with the Union of Concerned Scientists in Washington DC, says that he strongly supports exploring the storage of down-blended weapons-grade plutonium at WIPP. Such waste generates much less heat than does spent fuel, he adds. But he rejects storing spent fuel at WIPP, as its likely impacts on the surrounding salt "would be inviting trouble".

The DOE Field Office in Carlsbad and the Nuclear Waste Partnership had not responded to *Nature* when this article went to press.

Several scientists say that whatever the test results or arguments, the storage of high-level waste at WIPP should be ruled out because of the nature of the site. The area is rich in oil, gas and minerals, and oil and gas wells hug the 41-square-kilometre area. Hydraulic fracturing — fracking — of gas is also carried out nearby. This poses the risk that the WIPP repository could be disturbed by future drilling and mining, for example, by the puncture of the high-pressure brine reservoirs beneath WIPP.

There is no way that the authorities would ever approve such a site for storing high-level waste, says Chaturvedi. ■

DRUG DEVELOPMENT

Microbiome therapy gains market traction

Wave of investment suggests drugs from body-dwelling bacteria are heading for the clinic.

BY SARA REARDON

The human body teems with trillions of microorganisms — a microbial landscape that has attracted roughly US$500 million in research spending since 2008. Yet with a few exceptions, such as the use of faecal transplants for treating life-threatening gut infections or inflammatory bowel disease, research on the human microbiome has produced few therapies.

That is poised to change as large pharmaceutical companies eye the medical potential of manipulating interactions between humans and the bacteria that live in or on the body.

On 2 May, drug giant Pfizer announced plans to partner with Second Genome, a biotechnology firm in South San Francisco, California, to study the microbiomes of around 900 people, including those with metabolic disorders and a control group. "We are looking at using this as one piece of a puzzle to understand an individual," says Barbara Sosnowski, vice-president of external research and development at Pfizer in New York. A day earlier, Paris-based Enterome revealed that it had raised €10 million (US$13.8 million) in venture capital to develop tests that use the composition of gut bacteria to diagnose inflammatory and liver diseases.

Experts predict that the next few months will see a boom in such partnerships and investments, and that new microbiome-derived drugs and therapies will come to market within a few years.

Probiotics, or beneficial gut bacteria, have become a popular therapy in recent years. Television advertisements feature celebrities touting *Bifidobacterium*-laced yogurt, and consumers flock to buy pills that contain *Lactobacillus* to quell their gut disturbances and other ailments. But many physicians and scientists doubt the effectiveness of such remedies. "Probiotics may be relatively safe, but not particularly potent in terms of modifying diseases or symptoms," says Joseph Murray, a gastroenterologist at the Mayo Clinic in Rochester, Minnesota.

EYE OF SCIENCE/SPL

Researchers are studying how gut bacteria such as *Lactobacillus* (grey) interact with the body.

But as scientists come to understand the mechanisms by which specific bacteria affect the body, many think that they can pinpoint the right combination of microbes to treat different conditions. Others aim to develop molecules that mimic a beneficial bacterium–host interaction, or block a harmful one. "Undoubtedly, the microbiome is a little drug factory in our intestine," says Justin Sonnenburg, a microbiologist at Stanford University in Palo Alto, California.

Murray's group, for example, has reported that feeding the gut bacterium *Prevotella histicola* to transgenic mice engineered to have human-like immune systems can suppress the inflammation caused by multiple sclerosis and rheumatoid arthritis. His team is hoping to develop this into a therapy with biotech firm Miomics in New York. Similarly, Vedanta Biosciences in Boston, Massachusetts, is conducting preclinical trials of a pill containing microbes that suppress gut inflammation (Y. Furusawa *et al. Nature* **504,** 446–450; 2013).

And last June, Second Genome announced a deal with Janssen Pharmaceuticals of Beerse, Belgium, to study the microbial populations of people with ulcerative colitis, in the hope of identifying new drugs and drug targets. Although Second Genome remains vague about the details of its products, president Peter DiLaura says that the company hopes to find small molecules and biological compounds such as proteins that can tweak the microbiome to ease diabetes and autoimmune disorders.

Meanwhile, one of Second Genome's scientific consultants, bioengineer Michael Fischbach of the University of California, San Francisco, is developing tools to identify molecules found on bacteria or produced by them, and which bind to receptors on human cells and affect the immune or nervous systems. "It's not just a drug-like molecule — it's a real drug being produced," he says.

Changing the balance of 'good' and 'bad' bacteria in the gut microbiome can also influence health — inflammation, for example, or even depression and anxiety. Researchers may already have a wealth of ready-made medications that can alter this equilibrium. Drugs and small molecules that have been discarded because they are not absorbed by the intestine may help to target the gut microbiome specifically, treating it as an organ.

MORE ONLINE

TOP NEWS

FRANK WALDECKER/LOOK DIE BILDAGENTUR DER FOTOGRAFEN/ALAMY

Ancient mountain range in Tibet pre-dated Himalayas go.nature.com/gkohw9

MORE NEWS

- More studies on male genitalia than on female parts go.nature.com/brfcmv
- Deepwater Horizon's legacy of methane go.nature.com/jugbn1
- Biomimetic veins deliver self-healing fluids go.nature.com/vscac8

NATURE PODCAST

Virgin males are rougher with pups; animal studies have gender bias; and droughts move mountains nature.com/nature/podcast

Sonnenburg's team, for instance, has found that a compound called sialic acid builds up in the intestine and helps harmful bacteria to take over the gut when antibiotics have killed off helpful bacteria. The researchers are now investigating whether treating mice with compounds similar to sialic acid can inhibit this harmful transformation (K. M. Ng *et al. Nature* **502,** 96–99; 2013).

And Microbiome Therapeutics, a biotechnology company in Broomfield, Colorado, is currently conducting clinical trials with two small molecules that select for 'good' gut bacteria to help people with diabetes to take up insulin more easily. Chief executive Steven Orndorff says that the company plans to present the first results from the trials next month at an Endocrine Society conference in Chicago, Illinois.

Other companies are turning the microbiome into a diagnostic tool. Enterome has created a genetic-sequencing platform that detects changes in stool microbes that warn of the onset of disorders such as inflammatory bowel disease. The firm has tracked progression of the disease in 100 such patients in a bid to avoid invasive colonoscopies.

Getting microbiome-inspired therapies to market presents a number of challenges, however. Small molecules such as those developed by Microbiome Therapeutics may be able to go through the normal drug regulatory pathway. But there may be a different or new set of regulatory hurdles for genetically modified bacteria — for example, those in development by Ghent-based ActoGeniX in Belgium and ViThera Pharmaceuticals in Cambridge, Massachusetts — that deliver anti-inflammatory agents to the gut. Other issues, including intellectual-property rights for naturally occurring bacteria, may complicate the path of products to market.

Although small start-up firms can be flexible in navigating these issues, funding and guidance from pharmaceutical giants can only help, says Bernat Olle, chief operating officer of Vedanta. In 2013, for example, Vedanta struck a deal with Johnson & Johnson, based in New Brunswick, New Jersey, to develop potential therapies for inflammatory bowel disease and other autoimmune disorders.

"The microbiome is a little drug factory in our intestine."

Pierre Belichard, Enterome's chief executive, says that such investment has been a long time coming — but companies are now flocking to microbiome research. "Doctors have been asking questions about why this new and fascinating world of science is not seen as a place to put money in," he says. "Until the beginning of this year, that was a very good question." Now, he says, investors "all want a microbiome company in their portfolio". ■

BEN MOAT

A sensor-equipped mooring that measures the strength of the Atlantic Ocean's overturning currents.

OCEANOGRAPHY

Atlantic current strength declines

But more data are needed to indicate whether the slowing is a result of human-induced climate change.

BY QUIRIN SCHIERMEIER

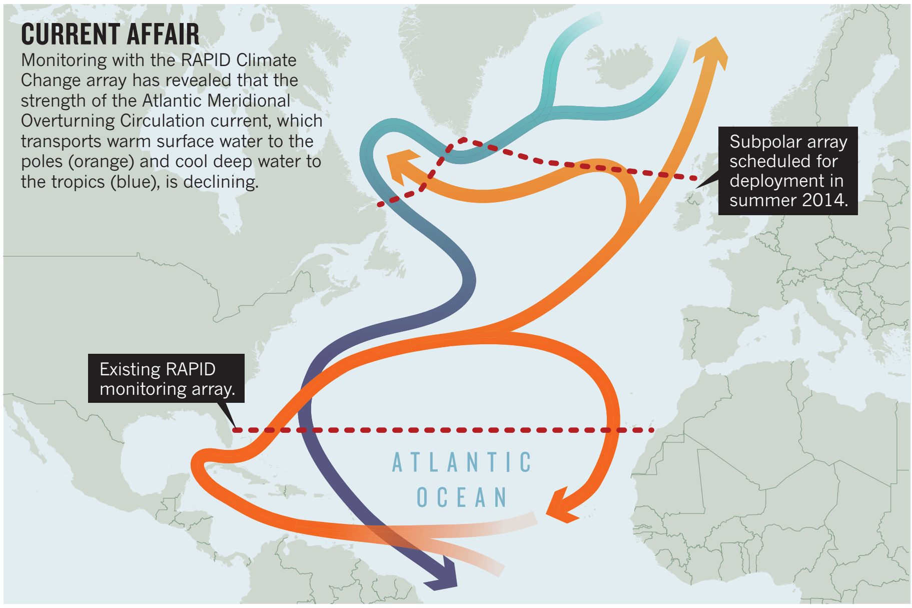
The marked slowdown in the past decade of the warm Atlantic Ocean currents that bring mild weather to northwestern Europe may be caused by natural variation and not anthropogenic climate change, as has been previously suggested.

The Atlantic Meridional Overturning Circulation (AMOC) is part of the great ocean 'conveyor belt' that ceaselessly circulates sea water, heat and nutrients around the globe. In particular, it transports large amounts of warm water from the tropics to the poles, warming the British Isles and maritime northern Europe along the way (see 'Current affair'). But since 2004, ocean sensors have detected a significant decline in the strength of the currents[1] and a cooling of the subtropical Atlantic as a result[2]. From mid-2009 to mid-2010, for example, the circulation slowed to two-thirds of its usual strength — and some oceanographers suggested that the drop caused the harsh weather in the United Kingdom and western Europe that winter (see *Nature* **497,** 167–168; 2013).

Climate scientists had speculated that the slowdown is linked to man-made climate change. But an analysis presented last month by a team of British scientists at the annual assembly of the European Geosciences Union in Vienna suggests that the AMOC's slowing could just be part of natural oceanic fluctuations. The researchers added, however, that it will take more long-term monitoring to definitively rule out climate change as a factor.

Scientists think that the AMOC might be subject to abrupt changes that have probably played a part in ancient climate events, such as the sudden temperature swings 18,000 to 80,000 years ago during the last glacial period. The AMOC's main engine — the sinking of cold, dense water to the bottom of the North Atlantic — has been identified as a potential 'tipping element' in Earth's climate system, in which small climate perturbations could push the system past a critical threshold, with potentially large consequences for humans and ecosystems[3].

Since 2004, 22 moored sensors have been deployed between the Canary Islands and



Florida along the latitude line at 26.5° north — where the AMOC emits its maximum heat. The sensor array, known as the RAPID Climate Change monitoring array, has continuously monitored the strength and temperature of the current at different depths.

RAPID measurements previously revealed[1] that the circulation weakened by 3% per year on average between 2004 and 2008, with a mean strength of 17.5 million cubic metres per second. Most of the past decade's observed decline occurred between April 2008 and March 2012, when the AMOC was around 15% weaker on average than in the previous four years. The measurements also showed that the strength of the currents varied by up to 70% from year to year, depending on wind and seawater temperature.

To find out whether the observed long-term decline lies within the range of natural yearly fluctuations, Chris Roberts, a climate scientist at the UK Met Office's Hadley Centre in Exeter who led the latest analysis, compared the observed trend with estimates of circulation strength derived from 14 state-of-the-art climate–ocean models. If the variability in modelled circulation strength were to differ substantially from observed trends, it could suggest that the decline is down to an external forcing factor such as climate change.

"It could have significant consequences for society."

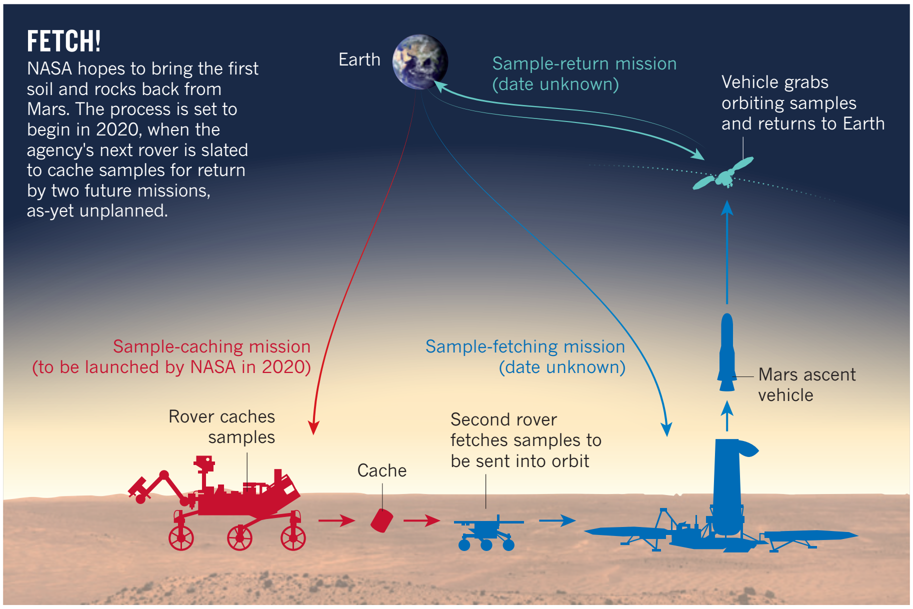
Although the results suggested that the downward trend is extremely unusual, Roberts knew that models can substantially underestimate the actual year-to-year variability in the strength of the AMOC. When he and his team adjusted the models to incorporate more-realistic natural fluctuations, the downward trend was statistically in line with the expected variations. Even if the slowing continues at the current rate, the trend will not differ significantly from plausible estimates of natural variability for 18 more years, the team concluded. But it will take at least 10 more years of continuous observation to detect any influence of man-made climate-change effects, says Roberts.

"There's nothing at the moment that would suggest that something dramatically worrying is going on," says David Smeed, an oceanographer at the UK National Oceanography Centre in Southampton and a lead researcher in the RAPID programme. He suggests that the weakening of the AMOC could be because of the Atlantic Multidecadal Oscillation — a natural cycle of ocean variability in which Atlantic temperatures dip every 60 to 70 years.

RAPID, which was funded by the Natural Environment Research Council in Swindon, UK, was last year extended to run until 2020. Another array, funded mainly by UK and US science agencies, will be deployed this summer in the North Atlantic between Labrador, Greenland and Scotland to monitor the AMOC in subpolar regions. Together, data from the two arrays should help to explain the mechanisms behind the changes in circulation, says Susan Lozier, an oceanographer at Duke University in Durham, North Carolina, especially because the subpolar array is along a similar latitude to the main driver for the Atlantic Ocean circulation system.

Regardless of the cause of the AMOC's decline, if the trend persists "it could have significant consequences for society" in terms of the climate in northwestern Europe, says Roberts. Nevertheless, being able to predict the strength of the current could help to improve short-term regional climate forecasts, he says. ■

1. Smeed, D. *et al. Ocean Sci.* **10,** 29–38 (2014).
2. Cunningham, S. A. *et al. Geophys. Res. Lett.* **40,** 6202–6207 (2013).
3. Lenton, T. M. *et al. Proc. Natl Acad. Sci. USA* **105,** 1786–1793 (2008).



IMAGES OF EARTH AND MARS SURFACE: NASA

PLANETARY SCIENCE

NASA plans Mars sample-return rover

Agency to narrow down list of landing sites for 2020 mission.

BY ALEXANDRA WITZE

NASA's Curiosity rover is in the prime of its life, exploring the rocks, soil and air of Mars. But the agency is already planning its successor — and this time, the scientific stakes are higher.

On 14 May, planetary geologists will gather in a hotel near Arlington, Virginia, to begin hammering out where NASA might send its next Mars rover, set to launch in 2020. The plan is to build a machine that is nearly identical to Curiosity, and equip it with fresh instruments to probe the Martian surface.

Although NASA has yet to finalize details, the next rover will almost certainly have a hugely important, unprecedented job: to collect and store rocks and soil for a future spacecraft to bring back to Earth. It would be the first ever sample return from Mars.

"The next 20 years of Mars exploration hinges on where this rover goes," says Philip Christensen, a planetary scientist at Arizona State University in Tempe. "It has to tell us something fundamental about the broader history of Mars."

NASA's workshop this week will discuss possible landing sites. Many look familiar: they were on the longlist of sites for Curiosity's landing in 2012. Such locations include Mawrth Vallis, an ancient valley strewn with minerals formed in water, which would help with the rover's main goal of finding and exploring environments that could once have been suitable for life. The European Space Agency is also considering the site for its ExoMars rover, which will launch in 2018 (see *Nature* **508,** 19–20; 2014).

Other possibilities for 2020 include several ancient, now-dry lakes and deltas where flowing water once laid down sediment. These areas, including Eberswalde Crater, were among the top candidates for the Curiosity mission. They were passed over in favour of Gale Crater, where the rover is laboriously trekking towards a 5-kilometre-high mountain of sediments. Curiosity has yet to detect concentrated amounts of organic material, but the rich river-laid sediments in Eberswalde are likely to offer that bounty, says geologist Ross Irwin of the Smithsonian Institution in Washington DC.

The 2020 rover will also have the crucial extra task of collecting samples. Scientists have talked for decades about getting their hands on Martian rocks to look for signs of past life. They have studied meteorites that originated on Mars, but no space agency has yet been able to bring back samples directly, in part because of the cost and in part because of technical failures (see *Nature* **479,** 275–276; 2011).

NASA's plan for bringing back Martian samples would involve a succession of missions over many years (see 'Fetch!'). Step one would need a rover to collect and store roughly 30 narrow cylinders of rock and soil, either on board or on the ground. In step two, an unmanned rocket would fly to Mars and deploy another rover to fetch the samples and then blast them into orbit. Step three would be to capture that orbiting package and fly it back to Earth.

Being able to look at a chunk of rock from a particular location and understand its context would be a crucial step forward, says John Mustard, a planetary geologist at Brown University in Providence, Rhode Island. "If samples were returned, the science that could come out of that would be equivalent to when the Apollo samples came back from the Moon," he says. "It changes everything."

Some scientists hope that the 2020 rover will visit new locations, whereas others want to return to sites explored by previous rovers such as Curiosity or Spirit. Both approaches have benefits, says Dawn Sumner, a geologist at the University of California, Davis. "With a new site there are more unknowns, but that means we are likely to learn more about the diversity of Mars as a planet," she says.

Much depends on what instruments the rover will carry. Fifty-eight teams have submitted proposals; in July or later, NASA will select a handful of designs. Curiosity's tools include an instrument that it deployed on 5 May to drill into a sandstone rock named Windjana. The rover has already drilled at two other sites, which yielded signs that an ancient lake bed once existed at Gale Crater.

The destination of the 2020 rover will not be constrained by science alone: engineers must be able to manoeuvre the craft to the ground safely. They may use a variant of Curiosity's 'sky crane' descent, in which thrusters guided the rover to a precise location.

This week's workshop is the first step in narrowing down the list of landing sites, but the final decision might not be made until 2019. "It will be a tremendously interesting and fascinating time," says Matthew Golombek, a planetary geologist at the Jet Propulsion Laboratory in Pasadena, California, who is leading the site-selection process. "The most important thing for this spacecraft is not so much to learn about the rocks on Mars, but to learn enough to know if those rocks have the stuff in them that you want to bring back to Earth." ■

CORRECTION

In the News story 'Doubts over heart stem-cell therapy' (*Nature* **509,** 15–16; 2014), the low-oxygen method for preparing mesenchymal stem cells was erroneously attributed to a Moscow research institute. The Moscow institute developed the experimental concept, but the low-oxygen refinement was developed in the United States. Mention of the institute has therefore been removed online.

DAI KUROKAWA/EPA/CORBIS

Kenyan activists in Nairobi protest against Uganda's anti-gay law.

Under siege

A wave of anti-gay laws and homophobia in Africa is hampering efforts to study and curb the spread of HIV.

BY LINDA NORDLING

On the morning of Saturday 12 April, ten police officers raided Maaygo, a men's health and HIV/AIDS advocacy organization in a residential area of Kisumu in western Kenya. Staff watched helplessly as the officers confiscated information leaflets and even the model penis used in condom demonstrations. The police arrested the organization's director and finance officer, as well as one of its members, for "illegally practising sexual orientation information".

The detainees were released later that day after Daniel Onyango, director of a local homosexual-rights group, arrived at the police station and explained that it was not actually a crime to distribute information on sexual orientation. Maaygo will continue its operations — albeit from a less conspicuous location in Kisumu.

An HIV treatment and research project in Uganda was not so lucky. On 3 April, the Makerere University Walter Reed Project in Kampala suspended its operations indefinitely after a staff member was arrested on charges of recruiting homosexuals and carrying out 'unethical research'.

Stories such as these are becoming increasingly common across Africa, where vocal protests against homosexuality are on the rise. Many countries, including Ethiopia, Nigeria and Senegal, have stringent anti-homosexuality laws. In February, Uganda toughened its stance, passing harsh new measures that hand down life sentences to people convicted of having gay sex and that stipulate up to seven years in jail for actively supporting the rights of gay people.

Antipathy towards homosexuality has hampered efforts to curb HIV in these countries. The Joint United Nations Programme on HIV/AIDS (UNAIDS) has identified men who have sex with men (MSM) as a key risk group for HIV infection, but because of cultural prejudices, gay people in Africa are often unable to access information on how to protect themselves from HIV, and those who become infected are often denied treatment. Homophobia and criminalization are also impeding research on MSM and HIV transmission.

"There are several examples where research has been stopped or slowed based on these laws," says Stefan Baral, an epidemiologist and physician at Johns Hopkins Bloomberg School of Public Health in Baltimore, Maryland. He has seen the problem at first hand, having conducted research with MSM in several African countries over the past decade. "What you end up with then is a data paradox, where you

know the least in the places with the most stigma," he says.

Researchers and clinicians say that there are a few small signs of hope. In Kenya, for example, one clinic has managed to use education to curb anti-gay sentiments, providing a model that might be successful elsewhere. But that relative stability could easily crumble with changes in the local community or in the country's leadership.

FORBIDDEN RESEARCH

The tide of homophobia has hit particularly hard in Uganda, says Paul Semugoma, a gay physician and researcher from that country who left two years ago and now lives in exile in South Africa.

"Uganda is the worst, it's a witch-hunt," he says. A research project was supposed to start there this year to identify groups at risk of HIV, but it was suspended because of concerns for the safety of the researchers and participants, he says. When it comes to MSM, he adds, "what's going to happen is that there's not going to be any up-to-date information on this risk group".

Similar problems are plaguing research in Ethiopia, where same-sex encounters are punishable by up to 15 years in prison. Researchers are kept from studying MSM and HIV by the Ethiopian Public Health Institute, which must approve medical research in the country.

A programme run by the US Centers for Disease Control and Prevention and the Ethiopian Public Health Association managed to pass the screening process in 2011 because it used terms such as 'most at-risk populations' rather than MSM or gay, says an Ethiopian advocate for gay and transgender health and human rights, who lives in exile in the United States and asked not to be named because of concerns about the safety of his family and friends. Once the government found out that the project would target MSM and related groups, the research was stopped, he says.

The problem is not limited to countries where homosexuality is a crime. "In many countries in sub-Saharan Africa, we just cannot do research on MSM-related topics," says Lung Vu, an HIV and tuberculosis research adviser at Population Services International, a global-health organization in Washington DC. "Government officials and religious leaders just don't allow for this to happen."

This is detrimental, he says, because data on MSM are needed to understand and combat HIV in Africa, where there is an epidemic of the disease. In an analysis[1] last year, Vu and his colleagues found that the prevalence of HIV among MSM in three large Nigerian cities was between four and ten times that in the general population. Scraps of data from Senegal and other countries paint the same picture.

Because there is so little information on MSM in Africa, researchers are only now starting to appreciate how much that population and other sexual minorities are spurring the spread of HIV on the continent, says Kent Klindera, director of an initiative at amfAR, the Foundation for AIDS Research in New York, that promotes HIV research and prevention for MSM and transgender individuals.

Funding problems have also limited research in this area, says a 2013 amfAR report[2]. Of the US\$1.5 billion allocated to six African countries (Botswana, Malawi, Namibia, Swaziland, Zambia and Zimbabwe) by the Global Fund to fight AIDS, Tuberculosis and Malaria since 2001, only 0.07% went to research concerning MSM and transgender people. Four of the countries reported no dedicated MSM projects whatsoever.

Despite the problems, researchers in Africa point to a few bright spots. For example, Mtwapa, a coastal town north of Mombasa, Kenya, managed to calm a volatile situation through community engagement and education about how the research can help society at large.

The trouble at Mtwapa centred on an HIV clinic run by the Kenyan Medical Research Institute (KEMRI), which conducted risk-group studies at the facility. On 12 February 2010, a mob of several hundred people charged the clinic, incited by two religious leaders — a Christian bishop and a Muslim imam.

➲ NATURE.COM
For a *Nature* special on research in Africa, see:
nature.com/africa

The riot was based on misinformation. "It started with a rumour that two gay men were getting married in the town," says Eduard Sanders, an epidemiologist with the University of Oxford, UK, who has studied MSM in Mtwapa since 2005, and who witnessed the riot. "But when the mob couldn't find any hint of the wedding, it descended on the clinic because of its well-known research on MSM."

The problem isn't limited to countries where homosexuality is a crime.

Armed with sticks, stones and other weapons, the crowd surrounded the clinic, demanding that the gay men come out. Police arrested people accused of being gay — possibly as a way of saving them from mob justice — and later released them. One KEMRI volunteer was severely beaten, according to the international group Human Rights Watch.

ATTITUDE ADJUSTMENT

Today, the clinic is safer, says Sanders, thanks to a campaign to inform local residents about its role in managing HIV in the area. As part of this effort, KEMRI hired community-liaison officer Evanson Gichuru, who met and interviewed local leaders and the people who organized the attack. Most of the antipathy, he observed, came down to misconceptions about gay people, such as the idea that they groom young boys to become prostitutes. Gichuru's meetings, combined with a media campaign and health-worker training on MSM sensitivities, have brought a turnaround in attitudes. There have been no further attacks on the clinic, and research conducted after health-worker training found[3] that it significantly reduced homophobic tendencies, particularly among those who had previously scored a high rating on a homophobia scale.

An important factor in changing attitudes was teaching people that the health of the wider community in Kenya depends on understanding and treating HIV among MSM, says Gichuru. Many men who have sex with men in Kenya also have sex with women, so a high HIV incidence in this group puts society in general at risk.

But neither Sanders nor Gichuru wants to overstate the success in Mtwapa. "I don't want to be overconfident," says Sanders. There is still a big stigma to being gay in Kenya, he adds.

And the techniques used to raise awareness there are not likely to work in Uganda, says Semugoma. "In Uganda the homophobia is not only from the people, it's also coming from the state," he says.

Researchers point to Malawi as another example of positive developments. A few years ago, the Centre for the Development of People (CEDEP), a human-rights organization that supports minorities including MSM, had to scale down its activities after health workers were arrested. But in 2012, partly thanks to CEDEP's advocacy, Malawian President Joyce Banda suspended the country's laws criminalizing homosexuality. The arrests have stopped and CEDEP now has five offices across Malawi.

The raid on Maaygo's office in Kenya last month worried many physicians and researchers, but it also provided an opportunity. In subsequent meetings with local authorities and the police, people from the organization and other human-rights groups discussed how HIV treatment and support for MSM are needed to fight the high HIV infection rates in Kisumu, says Onyango, who attended the first such meeting. "Currently all is well. We have not had any further incidents," he says. "Maaygo are still doing their activities, although at a slow pace so that we can sensitize the community." ■

Linda Nordling *is a freelance writer in Cape Town, South Africa.*

1. Vu, L. *et al. J. Acquir. Immune Defic. Syndr.* **63,** 221–227 (2013).
2. amfAR & Johns Hopkins Bloomberg School of Public Health *Achieving an AIDS-Free Generation for Gay Men and Other MSM in Southern Africa* (2013).
3. van der Elst, E. M. *et al. J. Int. AIDS Soc.* **16** (Suppl. 3), 18748 (2013).

FIRST LIGHT

The left-over radiation from the Big Bang has given up what may be its last great secret about the early Universe, but astronomers are determined to mine more from this primordial prize.

by Joanne Baker

Cosmologists couldn't have wished for a better anniversary present. Almost 50 years to the day after the first detection of the Big Bang's afterglow — a faint glimmer of long-wavelength photons known as the cosmic microwave background (CMB) — the field has been galvanized by what may be the last major discovery from the radiation.

On 17 March, astronomers announced that a microwave detector at the South Pole had recorded the first signs of primordial 'B modes': subtle, swirling patterns in the CMB data that were imprinted during the early history of the Universe. The result was hailed as direct evidence of gravitational waves, ripples in the fabric of space-time that were produced by a sudden 'inflation' of the Universe a split second after the Big Bang.

The most obvious question — is the B-mode signal real? — has sparked a race among teams running telescopes on the ground, in space and carried by balloons. "The name of the game is confirmation," says experimental cosmologist Amber Miller of Columbia University in New York City.

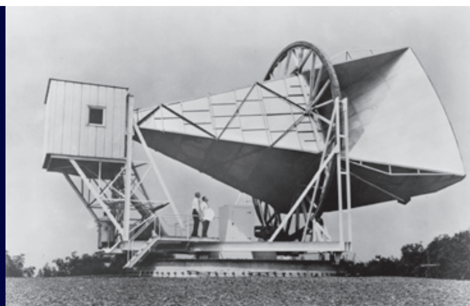
Should the results check out, and most in the field think they will, the focus will shift to the next frontier. Scientists would like to see a new era of B-mode astronomy that would collect more extensive and more precise measurements of the patterns. Through such data, researchers hope to reach back in time to better understand the Universe's first moments, as well as how galaxies formed and clustered together in the aftermath of the Big Bang. The B-mode data may even help to reveal the origins of mysterious factors such as dark matter and dark energy that control the form and fate of the cosmos.

"The CMB has been our best window on the early Universe by a long shot," says George Efstathiou, a cosmologist at the University of Cambridge, UK. But a rich new B-mode era is not guaranteed. Funding is scant, the existing surveys have little coordination with each other, and the available instruments are limited. What's more, theorists still need to pin down exactly what the new views of the CMB can reveal. Even as

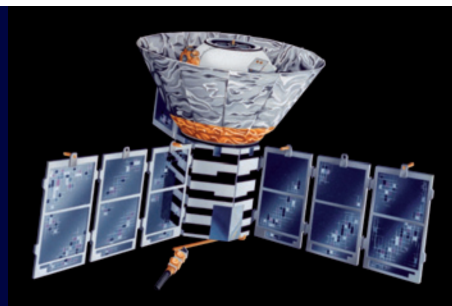
1964, NASA; 1990, NASA/COBE SCIENCE TEAM; 1992, NASA

50 YEARS OF DISCOVERY

Half a century after astronomers first detected the cosmic microwave background (CMB) radiation, it continues to be their clearest window on the early Universe.



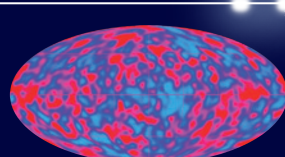
1964
Arno Penzias and Robert Wilson detect the CMB radiation and measure its temperature to be roughly 3 kelvin.



1990
NASA's Cosmic Background Explorer (COBE) satellite measures the CMB from space and pins its temperature at 2.725 kelvin.

1946–1948
Several scientists predict that the Universe should be filled with remnant radiation from the Big Bang, and that this would have a temperature of just a few kelvin.

1992
COBE data reveal minuscule variations in the CMB's temperature, a sign of density fluctuations in the early Universe that would later condense into galaxies.



they celebrate this year's discovery, CMB researchers are fretting over the future of their field. Decisions made over the next few months will determine whether astronomers can hope to realize the scientific promise of this new vista in the next decade or more.

EARLY DAYS

The discovery of the Big Bang's afterglow came as a happy accident, when Arno Penzias and Robert Wilson, two astronomers at Bell Labs in Holmdel, New Jersey, set out to map radio emissions from the Milky Way. On 20 May 1964, they noticed a faint signal that seemed to come from every direction. Penzias and Wilson assumed it was an artefact from some local source until a conversation with a colleague led them to conclude that the radiation was not earthly but cosmic.

Theorists, they learned, had long predicted such a signal: it was strong evidence in favour of the Big Bang theory, which holds that the Universe exploded into existence at a moment in the past, rather than having existed forever in an unchanging 'steady state'. By observing it, Penzias and Wilson had proved that the Universe was once much hotter than it is today. The photons they had recorded were released about 380,000 years after the Big Bang, when the expanding cosmic fireball had cooled enough for electrons and protons to combine to form hydrogen atoms. The photons have been travelling ever since, cooling as the Universe expands and preserving a snapshot of the Universe at the moment they were liberated (see '50 years of discovery').

In 1990, NASA's Cosmic Background Explorer (COBE) satellite made the first precise measurement of the CMB's temperature — 2.725 kelvin — and showed that the value was the same in every direction, implying that the primordial plasma was similarly uniform[1].

But it soon became apparent that the CMB is not perfectly smooth. In 1992, COBE scientists found that the temperature of the CMB varies across the sky by roughly 1 part in 100,000 (ref. 2). These tiny 'anisotropies' turned out to provide crucial information about the evolution of the Universe. The hot and cold spots reflect small variations in the density of the gas when the CMB photons were released. Most cosmologists think that gravity later magnified these fluctuations, pulling together denser regions to form galaxies and clusters of galaxies.

Seeing the anisotropies also inspired theorists, says Marc Kamionkowski, a cosmologist at Johns Hopkins University in Baltimore, Maryland. A prime example was the recognition that the warm and cold blotches in the CMB have characteristic sizes determined by vast waves of pressure and density that reverberated through the infant cosmos in much the same way that sound harmonics echo inside a violin. These dominant frequencies, or acoustic peaks, in the CMB allow astronomers to infer many physical properties of the Universe. For example, the biggest peak, akin to the loudest harmonic, lies at a scale of about 1°, or about twice the diameter of the full Moon. This is exactly as would be expected if the expanding Universe is geometrically flat, so that parallel light rays never cross as they traverse space. The location and relative strength of the second peak, at roughly 0.4°, allows astronomers to infer that ordinary matter — the kind found in atoms, planets and stars — comprises less than 5% of the cosmic total. Everything else is in the form of invisible dark matter and dark energy.

POLARIZATION PATTERNS

CMB research entered a new phase a decade ago, with the advent of detectors sensitive enough to measure its polarization — the direction of vibration in the photons coming from each point in the sky. Polarization in the CMB results from photons scattering off free-roaming electrons in the cosmic plasma, and the potential scientific pay-off from measuring it was huge: one component, the swirling B modes, promised to give astronomers the first direct evidence that the Universe had undergone an extreme form of inflation when it was just 10^{-36} to 10^{-32} seconds old. Theorists proposed the idea in the early 1980s to explain why the Universe is both smooth at the largest scales and geometrically flat[3]. The rapid expansion, in which the cosmos grew by a factor of at least 10^{26}, would have smoothed out most irregularities and flattened out any curvature. The few remaining irregularities — visible as the CMB temperature anisotropies — were vastly magnified vestiges of tiny quantum fluctuations in energy.

But that was all theory until researchers developed the capacity to measure B modes. That required them to find a way to identify a minuscule signal that is easily masked by polarized emissions from dust and magnetic fields in our Galaxy. The first detection wasn't announced until 2013 (refs 4, 5) — and even then, the measurements were made on a small angular scale, at which polarization patterns are distorted by the gravitational fields of galaxies in front of the CMB.

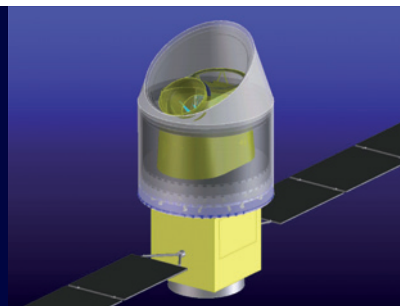
The real prize arrived in March, when astronomers working with the BICEP2 detector at the South Pole announced that they had measured B modes on scales of about 1°, large enough to avoid the signal from intervening galaxies and to probe fundamental polarization patterns such as those from inflationary gravitational waves[6].

After so many years of searching for inflationary B modes, BICEP2's results triggered widespread elation in the cosmology community. "It has injected a whole lot of adrenaline into the endeavour," says experimental cosmologist Shaul Hanany of the University of Minnesota in Minneapolis.

But with that excitement came a puzzle. The patterns detected by BICEP2 are considerably stronger than most cosmological models

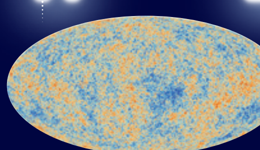
1999
Balloon-borne detectors characterize CMB fluctuations accurately enough for scientists to do a statistical analysis, which reveals information on the Universe's geometry and energy content.

2014
The BICEP2 experiment at the South Pole detects strong evidence of gravitational waves in the CMB's polarization.



2020s
Next-generation CMB observatories could use the radiation to track galaxy evolution and probe the earliest instants of the Universe.

2003
NASA's Wilkinson Microwave Anisotropy Probe (WMAP) charts the CMB in increased detail.



2013
Europe's Planck satellite picks up first hints of gravitational waves from the infant Universe.

1999, BOOMERANG/NASA/NSF; 2013, ESA/PLANCK COLLABORATION

2014, STEFFEN RICHTER/HARVARD UNIV.; 2020S, LITEBIRD

predicted. And it exceeds limits set by the European Space Agency's now-deactivated Planck satellite on the degree to which gravitational waves might have contributed to the CMB temperature fluctuations. "The BICEP2 result was a bit of a shock for me," says Efstathiou, who is a member of the Planck science team. "I think the jury is still out" on what it means, or if it's even real.

In the next year, half a dozen experiments in Antarctica and Chile will try to confirm the findings. Members of the BICEP2 team are working on two new South Pole telescopes. The Keck Array, which is already operational, has five times as many detectors as BICEP2 and covers two frequency bands. A second, called BICEP3, is an upgraded version of the previous detector that is scheduled to start collecting data in December 2014. One or two US-funded balloon-borne experiments may also fly later this year from McMurdo Station in Antarctica; last year's flights were cancelled because of the US government shutdown.

But cosmologists are most eager to see this autumn's planned release of the full data set from Planck — findings that will include polarization maps. Planck has the advantage of monitoring a wider range of frequencies than ground- and balloon-based experiments, which can measure only within the narrow bands of radiation frequencies that are not absorbed by water vapour in the atmosphere. Planck's improved sight will give astronomers much more confidence in subtracting foreground polarization from our Galaxy. And instead of being restricted to the portion of sky visible from a given latitude, Planck has an unobstructed view.

If Planck confirms BICEP2's results, the champagne will come out. But if it does not, cosmologists will have to resolve the discrepancy, which will pose a challenge. To take just one example, gravitational waves as strong as those recorded by BICEP2 should have had a noticeable effect on the acoustic peaks — yet the limited Planck data currently available have shown no evidence for that.

"How can you reconcile it?" wonders Efstathiou. The ideas put forward so far seem contrived, he says. Kamionkowski is more optimistic, and counsels patience. "It may take years to really understand what the most promising models are and how to distinguish them."

"It may take years to really understand what the most promising models are and how to distinguish them."

NEXT-GENERATION EXPERIMENTS

In the meantime, most CMB scientists are focusing on developing their capability for measuring B modes. For example, there are many theories for precisely how inflation unfolded, each making its specific prediction about how the gravitational wave B modes are distributed across the sky. Being able to measure B modes at the largest scales would allow astronomers to weed out the theories that are obviously wrong.

At smaller scales, the B modes are sensitive to how mass is distributed around the Universe, and how vast galaxy clusters have grown over time. Such a signal would help astronomers to constrain stubborn cosmological unknowns, including the nature of dark energy — a mysterious force that is causing the Universe's expansion to speed up — and the identity of the invisible dark-matter particles that make up most of the Universe's mass.

Combining B-mode maps with surveys of hydrogen across the Universe could also allow observers to probe the epoch in which the first stars and galaxies switched on their ionizing radiation. Electron scattering from this period should have left a large-scale mark in the B-mode polarization of the CMB.

Unfortunately, limited money is constraining choices about what comes next. A UK competitor to BICEP2 was cancelled in 2009 (see *Nature* http://doi.org/fnsdc3; 2009) as its funding council struggled to meet commitments to international bodies such as CERN, Europe's particle-physics laboratory near Geneva in Switzerland. Europe as a whole, meanwhile, has focused its CMB research programme almost exclusively on Planck — a policy that Efstathiou describes as a "big mistake". With no follow-on mission in the pipeline and little ground-based activity, the concern is that hundreds of postdocs and students will have to move to new fields when the research programme ends.

In the United States, CMB work falls into the cracks between grant-agency panels. Space and balloon work competes against planetary and X-ray astronomy missions at NASA, and ground-based arrays go up against particle-physics experiments at the Department of Energy (DOE) and the National Science Foundation (NSF). Philanthropic donations from the Keck and Simons foundations, among others, are helping fill the gap.

A solution suggested by some astronomers would be to cut back on the number of ground-based CMB experiments with similar aims. Critics contend that ground-based CMB experiments rarely share their data, which undermines calls for more projects. Most space-science missions are required to make their data public, says Jean-Loup Puget, an astronomer at the University of Paris-South and a principal investigator on Planck. "Ground-based experiments should do so too," he says.

But others argue that the ground experiments are cheap and that diversity makes for a healthy field. The one thing that everyone agrees on is that the science case for another CMB space mission is compelling. Efforts are afoot to make that happen after 2020.

It may be an uphill battle. No CMB probe was ranked highly in NASA's 2010 Decadal Survey, a community-led review that sets scientific priorities to guide future mission selection. But there was a clause recommending a mid-decade review in the event that B modes were discovered. After BICEP2, the US CMB community, led by Hanany and Jamie Bock of the Jet Propulsion Laboratory in Pasadena, California, is urging that the case for such a mission be reappraised and funding be redirected from existing missions that have suffered delays.

A consortium of US experimenters is proposing a follow-up to the present South Pole and Atacama telescopes. Known as CMB-S4, it would have hundreds of thousands of detectors and could come online after 2020 if it is given a high priority in a particle-physics review currently being carried out by the DOE and the NSF. Balloons could also play a part. "A coherent programme is warranted," Hanany says.

In Europe, a higher-resolution successor to Planck has not so far been selected by the European Space Agency. A revised mission is being drawn up for the next round of proposals, and if successful might be launched in the mid-2020s.

But such complicated proposals are expensive and difficult to realize, Efstathiou says. 'Keep it simple' is now his mantra. He would like to see a small mission dedicated to observing B modes on large angular scales, thus targeting the gravitational-wave signature alone. In effect, it would be a BICEP2 experiment in space, says Peter Ade at Cardiff University, UK, who has built detectors for ground-based experiments and Planck. The technology is mature and he thinks such a mission could be ready in five years.

A Japanese-led satellite proposal called LiteBIRD could be just such a mission. Proposed by physicists in Japan, in collaboration with experimenters in the United States, Germany and Canada, the project could be launched in the early 2020s if it received some US$100 million in funding. In the meantime, the researchers are developing a ground-based experimental version, called GroundBIRD.

For all the uncertainties about the future, CMB scientists are in bullish mood. "Nature has been kind to us in giving us this clear view of the early Universe," Efstathiou says. "With a gift like that we should exploit it as much as we can." ■

Joanne Baker *is a Comment editor for* Nature.

1. Mather, J. C. *et al. Astrophys. J.* **354,** L37–L40 (1990).
2. Smoot, G. F. *et al. Astrophys. J.* **396,** L1–L5 (1992).
3. Guth, A. H. *Phys. Rev. D* **23,** 347–356 (1981).
4. Hanson, D. *et al. Phys. Rev. Lett.* **111,** 141301 (2013).
5. The POLARBEAR Collaboration Preprint at: http://arxiv.org/abs/1403.2369 (2014).
6. BICEP2 Collaboration Preprint at: http://arxiv.org/abs/1403.3985 (2014).

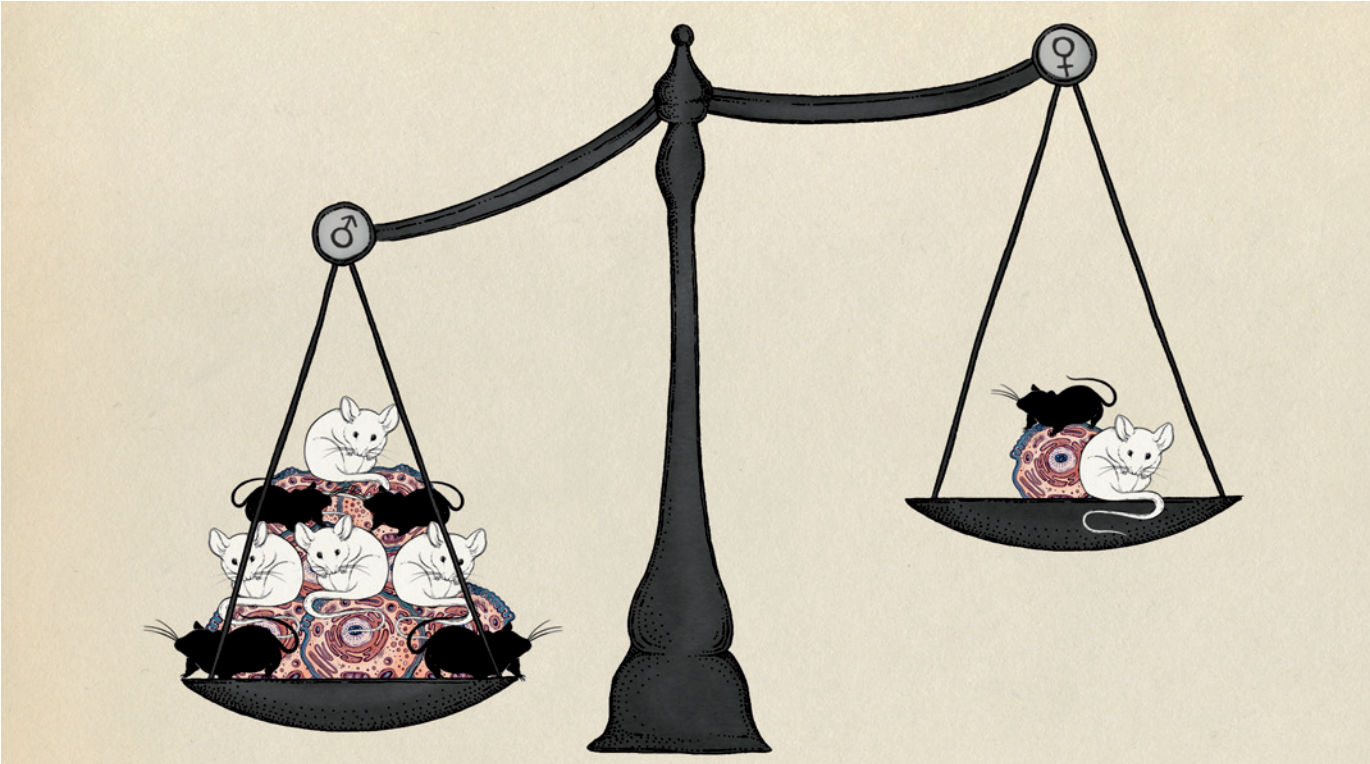


ILLUSTRATION BY KATIE SCOTT

NIH to balance sex in cell and animal studies

Janine A. Clayton and **Francis S. Collins** unveil policies to ensure that preclinical research funded by the US National Institutes of Health considers females and males.

More than two decades ago, the US National Institutes of Health (NIH) established the Office of Research on Women's Health (ORWH). At that time, the Congressional Caucus for Women's Issues, women's health advocacy groups and NIH scientists and leaders agreed that excluding women from clinical research was bad for women and bad for science. In 1993, the NIH Revitalization Act required the inclusion of women in NIH-funded clinical research.

Today, just over half of NIH-funded clinical-research participants are women. We know much more about the role of sex and gender in medicine, such as that low-dose aspirin has different preventive effects in women and men, and that drugs such as zolpidem, used to treat insomnia, require different dosing in women and men.

There has not been a corresponding revolution in experimental design and analyses in cell and animal research — despite multiple calls to action[1]. Publications often continue to neglect sex-based considerations and analyses in preclinical studies[2,3]. Reviewers, for the most part, are not attuned to this failure. The over-reliance on male animals and cells in preclinical research obscures key sex differences that could guide clinical studies. And it might be harmful: women experience higher rates of adverse drug reactions than men do[4]. Furthermore, inadequate inclusion of female cells and animals in experiments and inadequate analysis of data by sex may well contribute to the troubling rise of irreproducibility in preclinical biomedical research, which the NIH is now actively working to address[5,6].

The NIH plans to address the issue of sex and gender inclusion across biomedical research multi-dimensionally — through programme oversight, review and policy, as well as through collaboration with stakeholders including publishers. This move is essential, potentially very powerful and need not be difficult or costly.

➲ NATURE.COM
Read about NIH reproducibility policy at:
go.nature.com/rcr1ef

BETTER WITH BOTH

Certain rigorous studies evaluating the effects of sex differences have been effective in bridging the divide between animal and human work. One example concerns multiple sclerosis (MS). Women are more susceptible to MS than men are, but develop less-severe forms of the disease. The most widely accepted MS animal model — rodent experimental autoimmune encephalomyelitis (EAE) — has revealed[7] that sex differences in MS are related to both reproductive and non-reproductive factors. Findings[8] that oestrogen therapy provided benefits in rodent EAE supported use of an oestrogenic ligand as a candidate neuroprotective agent for MS that is now being studied.

Moreover, differences between the sexes in both the animal model and human MS have

now been correlated with genetic factors. For example, some Y-chromosome genes (in male mice) seem to have a protective effect against the disease, and some X-chromosome genes (in female mice, with potentially double the dosage) have a disease-causing effect. Earlier this year, a study[9] demonstrated that mice with XY chromosomes in the central nervous system had greater neurodegeneration than did those with XX chromosomes. The findings have important implications for other sex-skewed neurological conditions, including Parkinson's disease, schizophrenia and stroke. Finally, inherited effects have been linked to imprinting of genes on sex and non-sex chromosomes (autosomes). Maternal parent-of-origin effects have been associated with MS risk[10].

Substance abuse is also affected by sex. One target for intervention has been stress systems that mediate craving. Female rats exhibit a greater response to stress by the neurotransmitter norepinephrine than do male rats. A promising study[11] published this year provides the first evidence, in humans, of temporary attenuation of cocaine and alcohol craving, anxiety and negative emotion after stress in females — but not males — using guanfacine, which dampens the body's nervous-system response to stress.

Typically, reasons for male focus in animal-model selection centre on concerns about confounding contributions from the oestrous cycle. But for most applications, female mice tested throughout their hormone cycles display no more variability than males do, as confirmed in a meta-analysis[12].

Convention is another probable reason for reliance on the male-only models that have been typical in many research areas for decades. Lack of understanding about the potential magnitude of the effect of sex on the outcome being measured is likely to perpetuate this blind spot.

The sex of cell lines studied *in vitro* is also too often ignored. Female and male cells respond differently to chemical and microbial stressors. These intrinsic differences are hormone-independent but also exhibit further variation on differentiation and exposure to sex hormones. It is well known that many neurological conditions are sexually dimorphic, and cell-culture studies have demonstrated that male (XY) and female (XX) neurons respond differently to various stimuli. Male neurons are more sensitive to stress from reactive oxygen species and excitatory neurotransmitters; female neurons are more sensitive to some stimuli that prompt the programmed cell death known as apoptosis[13]. Data support distinct cell-death signalling in female and male neurons with potential applications in treatments for stroke, brain injury and other conditions.

There are several approaches to rigorous preclinical research with a focus on sex and gender[14]. One, the four-core-genotypes model, can identify and distinguish between the effects of genes and the effects of hormones. The four genotypes in this model are XX gonadal males or females, and XY gonadal males or females. Using this model has, for instance, demonstrated influence of the sex-chromosome complement as a cause of sex differences in obesity and metabolism. On a high-fat diet, mice with two X chromosomes gained more weight than XY mice did, regardless of gonadal sex, and also developed a fatty liver and elevated lipid and insulin levels. These differences are attributable to X-chromosome dosage rather than to Y-chromosome effects[15].

> ***"Convention is another probable reason for reliance on the male-only models."***

Various organizations have taken steps to increase awareness and address unconscious bias about the importance of sex and gender in biomedical research. Several journals now require authors to specify sex- and gender-related information. This includes stating the sex of animals used (or in the case of primary cells or cultures, the sex of the animal from which cells are derived) and that of human participants in published studies.

NIH STEPS

The NIH is now developing policies that require applicants to report their plans for the balance of male and female cells and animals in preclinical studies in all future applications, unless sex-specific inclusion is unwarranted, based on rigorously defined exceptions. These policies will be rolled out in phases beginning in October 2014, with parallel changes in review activities and requirements. Because our goal is to transform how science is done, the first step will be the development and delivery of training modules and detailed policy informed by ongoing data analysis. As part of its initiative to enhance rigour, the NIH plans to disseminate training on experimental design for NIH staff, trainees and grantees. Evaluation of sex differences will be included in these modules.

In 2013, the ORWH, which oversees the NIH-wide research agenda related to sex and gender influences, launched a programme that provides funding supplements to existing grants to add subjects, tissues or cells of the sex opposite to that used in the original grant, or to increase the power of a study to analyse for a sex or gender difference by adding more subjects of either sex to a sample that already includes both males and females. Although this strategy enables the NIH to capitalize on the value of current research investments, we expect that such a mechanism will no longer be needed once policies on sex influences are implemented for preclinical research.

The ORWH will continue to work with the US Food and Drug Administration to co-fund the Specialized Centers of Research on Sex Differences programme, which supports interdisciplinary collaborations on sex and gender influences in health, and bridges basic- and clinical-research approaches. This programme also facilitates training in sex and gender considerations in experimental design and analysis. The ORWH will leverage lessons learned from these centres.

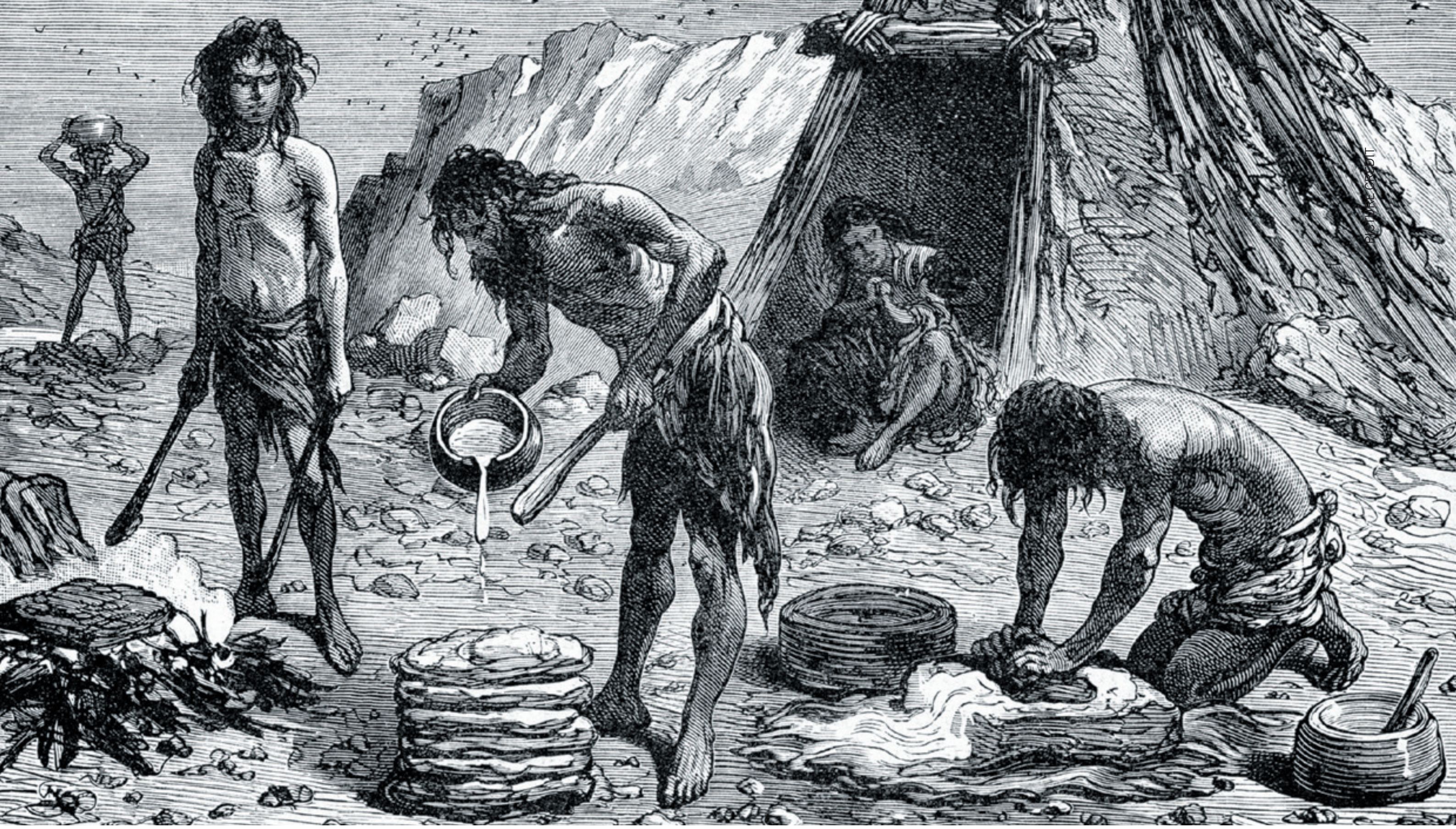
Reviewers of grant applications must also be brought to the table, because they provide the first insights into taxpayer-funded research. The NIH review process will be modified in phases, and coordinated with requirements for applicants. Reviewers will be enjoined to evaluate applicants' research plans to include, compare and contrast experimental findings in male and female animals and cells.

Furthermore, the NIH will monitor compliance of sex and gender inclusion in preclinical research funded by the agency through data-mining techniques that are currently being developed and implemented. Importantly, because the NIH cannot directly control the publication of sex and gender analyses performed in NIH-funded research, we will continue to partner with publishers to promote the publication of such research results.

In requiring sex and gender inclusion plans in preclinical research, the NIH will ensure that the health of the United States is being served by supporting science that meets the highest standards of rigour. ■

Janine A. Clayton *is director of the US National Institutes of Health Office of Research on Women's Health, and associate director for research on women's health, Bethesda, Maryland, USA.*
Francis S. Collins *is director of the US National Institutes of Health, Bethesda, Maryland, USA.*
e-mail: janine.clayton@nih.gov

1. Zucker, I. & Beery, A. K. *Nature* **465,** 690 (2010).
2. Beery, A. K. & Zucker, I. *Neurosci. Biobehav. Rev.* **35,** 565–572 (2011).
3. Mogil, J. S. & Chanda, M. L. *Pain* **117,** 1–5 (2005).
4. Franconi, F., Brunelleschi, S., Steardo, L. & Cuomo, V. *Pharmacol. Res.* **55,** 81–95 (2007).
5. Collins, F. S. & Tabak, L. A. *Nature* **505,** 612–613 (2014).
6. Landis, S. C. *et al. Nature* **490,** 187–191 (2012).
7. Voskuhl, R. R. & Palaszynski, K. *Neuroscientist* **7,** 258–270 (2001).
8. Wisdom, A. J., Cao, Y., Itoh, N., Spence, R. D. & Voskuhl, R. R. *J. Neurosci Res.* **91,** 901–908 (2013).
9. Du, S. *et al. Proc. Natl Acad. Sci. USA* **111,** 2806–2811 (2014).
10. Ebers, G. C. *et al. Lancet* **363,** 1773-1774 (2004).
11. Fox, H. C., Morgan, P. T. & Sinha, R. *Neuropsychopharmacol.* **39,** 1527–1537 (2014).
12. Prendergast, B. J., Onishi, K. G. & Zucker, I. *Neurosci. Biobehav. Rev.* **40,** 1–5 (2014).
13. Du, L., Bayir, H. & Lai, Y. *et al. J. Biol. Chem.* **279,** 38563–38570 (2004).
14. Becker, J. B. *et al. Endocrinology* **146,** 1650–1673 (2005).
15. Chen, X. *et al. PLoS Genet.* **8,** e1002709 (2012).



An artist's impression of the kind of social group in which early humans may have lived.

CCI ARCHIVES/SPL

SOCIOBIOLOGY

The distributed brain

Herbert Gintis salutes the follow-up to a study on sociality and hominin brain size.

Sociobiology was born in 1975, when biologist Edward O. Wilson published the volume that gave the field its name. There are many social species, and we can gain insight into human sociality by comparing it with sociality in other animals.

Several factors have propelled sociobiology to prominence in the behavioural sciences. Perhaps most important was the waning of a major impediment: the idea that linking human behaviour to genetics fuels racist ideology. Furthermore, as researchers learned about the lives of birds, primates and insects, the concept of social structure as a general biological category arose. Finally, the value of interdisciplinary research such as sociobiology became apparent in many areas in which complex systems cannot be understood using the conventional field categories.

Sociobiology in *Homo sapiens* is especially complex because of the enormous part played by culture in human evolution, which is explored in gene–culture coevolutionary theory. In *Thinking Big* — a follow-up to the edited volume *Social Brain, Distributed Mind* (Oxford University Press, 2010) — Robin Dunbar, Clive Gamble and John Gowlett explore the growth of the brain from early hominids to modern *Homo sapiens*, which has a ratio of brain to body mass three times that of other primates.

Thinking Big: How the Evolution of Social Life Shaped the Human Mind
CLIVE GAMBLE, JOHN GOWLETT AND ROBIN DUNBAR
Thames and Hudson: 2014.

Large brains are very costly. Increased cranial capacity required the restructuring of the human birth canal and led to birth before the fetus is fully matured. This in turn led to prolonged and collective child-rearing. In the average adult human, the brain represents about 2% of body weight, but consumes about 20% of calories. What could the counterbalancing advantages of large brains be?

The conventional answer has been skilful tool use. We now know, however, that hominin brain growth preceded by more than half a million years the emergence of material culture — visual art, hafted tools, crafted containers and written language, all of which began to appear some 70,000 years ago. Richard Byrne and Andrew Whiten's *Machiavellian Intelligence* (Oxford University Press, 1988) insightfully shifted the focus from technical to social skills, suggesting that a sharp wit conferred fitness by enabling individuals to deceive and manipulate. In this view, the large brain is the product of an arms race that is a drain on species-level fitness. Given the intense competition among hominins to fill the hunter-gatherer niche, this theory of human hypercognition seems implausible: the energy wasted in mutual deception would reduce human fitness, favouring small-brained competitors.

In *Thinking Big*, Dunbar, Gamble and Gowlett supply a more credible theory with their "social brain hypothesis". They describe the major findings of the ambitious 7-year project 'Lucy to Language: The Archaeology of the Social Brain', which involved more than 30 researchers and 5 UK universities, and was backed by the British Academy, the national funding body for humanities and social sciences. The authors show that there is a strong correlation between relative neocortex volume and mean social-group size in monkeys, apes and humans. They attribute this to the fact that the complexity of group interactions increases with group size. A large brain gives individuals the means to forge strong social ties that enhance their personal fitness and the group's social cohesion; in particular, a large neocortex

supports a "theory of mind", whereby individuals can form mental representations of the beliefs and intentions of others. This enables them to enter into complex agreements and coalitions, and to track multiple social relationships through time and space.

Social Brain, Distributed Mind is rather more detailed. It brings together an array of archaeologists, anthropologists, geographers, psychologists, palaeontologists, historians and philosophers involved in Lucy to Language, who together construct a plausible "cognitive anthropology" that defends the social brain hypothesis, while exploring the idea that the human mind is not confined to individual brains, but lives in a social network of minds across which cognition is distributed. They thus handle the problem of brain growth long preceding the emergence of material culture.

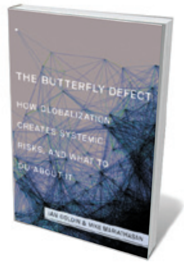
The contributors argue that long before material culture, early hominins developed a "material memory system" in the form of tokens and containers. This linked the mental power of many individuals, and led to a transition from knowledge acquisition and learning based on personal discovery to cognition based on social interaction and sharing. With distributed cognition, knowledge lies not only in the individual, but also in the social roles and iconic artefacts that link minds. For instance, humans have formed 'fission–fusion' social groupings in which kin relationships are maintained across subgroups, both males and females migrate to marry, and complex, powerful, fitness-enhancing familial alliances are sustained.

How plausible is the idea that we have big brains because we evolved to live in large groups, putting heavy cognitive demands on our ability to forge close social bonds with large numbers of individuals? My own view is that hunting required a high level of coordinated decision-making, and that the presence of lethal weapons undermined our ape ancestors' characteristic social-dominance hierarchy, which was based on the physical prowess of the alpha male. This created a leadership void that could be filled not by appeal to physical strength, but rather by social persuasiveness and then a subtle ability to form effective coalitions.

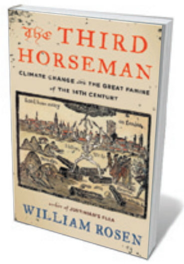
The result was a political structure in which linguistic facility and cognitive skills were rewarded with enhanced reproductive fitness. The social brain, then, helped our ancestors to operate successfully in a proto-democratic framework. ■

Herbert Gintis *is external professor at the Santa Fe Institute in New Mexico and professor of economics at Central European University in Budapest. His most recent book is* A Cooperative Species *(with Samuel Bowles).*
e-mail: hgintis@comcast.net

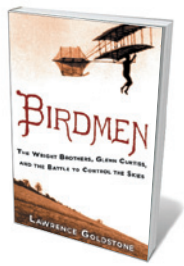
Books in brief



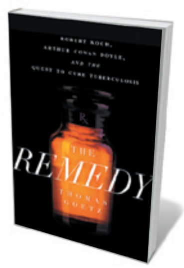
The Butterfly Defect: How Globalization Creates Systemic Risks, and What to Do about It
Ian Goldin and Mike Mariathasan PRINCETON UNIVERSITY PRESS *(2014)*
In a nod to chaos theory's butterfly effect — in which tiny perturbations unhinge big non-linear systems — this treatise explores globalization's built-in risks. Economists Ian Goldin and Mike Mariathasan analyse systemic vulnerabilities leading to cyber-attacks or pandemics, and look at the ecological risks integral to globalization. The sustainable management of such tangled interdependency, they argue, demands governance reform, including the setting up of research-led bodies to tackle big issues such as climate change.



The Third Horseman: Climate Change and the Great Famine of the 14th Century
William Rosen VIKING *(2014)*
A kink in Europe's climate during the fourteenth century indirectly triggered a seven-year cataclysm that left 6 million dead, William Rosen reveals in this rich interweaving of agronomy, meteorology, economics and history. The Great Famine ended the explosion in agricultural productivity of the 400-year Medieval Warm Period, which affected mainly North Atlantic civilizations. Rosen deftly delineates the backstory and the perfect storm of heavy rains, hard winters, livestock epidemics and war leading to the catastrophe.



Birdmen: The Wright Brothers, Glenn Curtiss, and the Battle to Control the Skies
Lawrence Goldstone BALLANTINE BOOKS *(2014)*
The daredevil scientists and engineers who forged the field of aeronautics spring vividly to life in Lawrence Goldstone's history. Wilbur Wright is famed for cracking the conundrum of powered, controlled, heavier-than-air flight through leaps of intuition and reasoning. Less known is his and brother Orville's feud with ace flyer and motor designer Glenn Curtiss. Goldstone never stints on the science in tracing the trio's patent wars and struggles to monopolize the industry over a decade of dazzling innovation.



The Remedy: Robert Koch, Arthur Conan Doyle, and the Quest to Cure Tuberculosis
Thomas Goetz GOTHAM BOOKS *(2014)*
What does germ theory have to do with evergreen fictional sleuth Sherlock Holmes? Science writer Thomas Goetz reveals all in this history of the hunt to cure tuberculosis (TB), centring on young physician Arthur Conan Doyle's 1890 trip to Berlin to report on bacteriologist Robert Koch's TB remedy, tuberculin. Conan Doyle rightly doubted its efficacy. But, impressed by Koch's postulates that particular organisms cause diseases, he intensified his focus on the scientific method and the hunting of other insidious villains in fiction.

Cold Blood: Adventures with Reptiles and Amphibians
Richard Kerridge CHATTO AND WINDUS *(2014)*
Nature writer Richard Kerridge fed, as a child, on accounts of black rhinoceroses, red river hogs and mandrills. His native Britain lacked such faunal glories — or so he thought, until he discovered the glistening hordes of amphibians and reptiles lurking in grass, bogs and leaf litter. In this mix of natural history, memoir and thoughts on the "cultural functions of wild animals for human beings", captured moments such as the golden flash of a palmate newt delight the reader as much as they did Kerridge's childhood self. **Barbara Kiser**

CHEMISTRY

Intoxicating science

Jamie Goode drinks in two views of that most venerable and destructive drug — alcohol.

Proof: The Science of Booze
ADAM ROGERS
Houghton Mifflin Harcourt: 2014.

The Drunken Monkey: Why We Drink and Abuse Alcohol
ROBERT DUDLEY
University of California Press: 2014.

"Booze is civilization in a glass," states Adam Rogers in *Proof*. This science-steeped tale of humanity's 10,000-year love affair with alcohol is an engaging trawl through fermentation, distillation, perception of taste and smell, and the biological responses of humans to booze. Robert Dudley's *The Drunken Monkey*, by contrast, focuses on the single question of why we drink — in many cases, to excess.

Of the two, *Proof* is the easier read. Rogers, a senior editor at *Wired* magazine, reveals how alcohol is a spin-off from a form of warfare: yeasts use it as a chemical weapon in their competition with other microbes. The sugar-rich environment of ripe fruits is a tantalizing food source for organisms ranging from bacteria to primates. So much so that yeasts use the relatively inefficient process of fermentation to metabolize the sugar, because it produces the waste product ethanol, which poisons competitors.

Humans first consciously exploited this around ten millennia ago: the oldest archaeological evidence of alcohol production is a pot shard dated to that time, from Jiahu in China. For some 9,850 of these years, fermentation must have seemed a mysterious, even mystical transformation — until, in 1857, French microbiologist Louis Pasteur revealed that yeasts are responsible. Even now, many aspects of booze — such as the link between soil characteristics and noticeable local flavours in wine — are yet to be fully explained. Exploring some of these unanswered questions is where Rogers has his fun.

Along with yeast, Rogers looks at the varied sugar sources that fuel booze production: rice, grapes and barley. The latter is generally malted, a process in which it is allowed to germinate a little to initiate the conversion of starch to sugar. Here we meet the remarkable Jokichi Takamine, a Japanese chemist who worked intermittently in the United States, and who in the late nineteenth century devised a way to break down starch without malting. His technique could have revolutionized whisky production, but it was never developed commercially because it threatened the livelihoods of maltmen. Rogers then takes us through fermentation and examines distillation, first invented some 2,000 years ago. Steam distillation — the process of heating and cooling through which alcohol and water, which have different boiling points, are separated — was probably invented in China, as evidenced by Han Dynasty bronze stills (a detail noted in *The Drunken Monkey*). Today most stills are copper, Rogers shows in *Proof*, because reactions with the metal get rid of the smelly volatile sulphur compounds produced by fermenting yeasts. As a result, the still's walls gradually thin, giving the vessel a lifespan of just 25 years or so.

HOLLER, HENDRIK/THE FOOD PASSIONATES/CORBIS

Fermenting fruit has held allure for millennia.

➲ NATURE.COM
For Harold McGee on fermented food, see:
go.nature.com/ysoefj

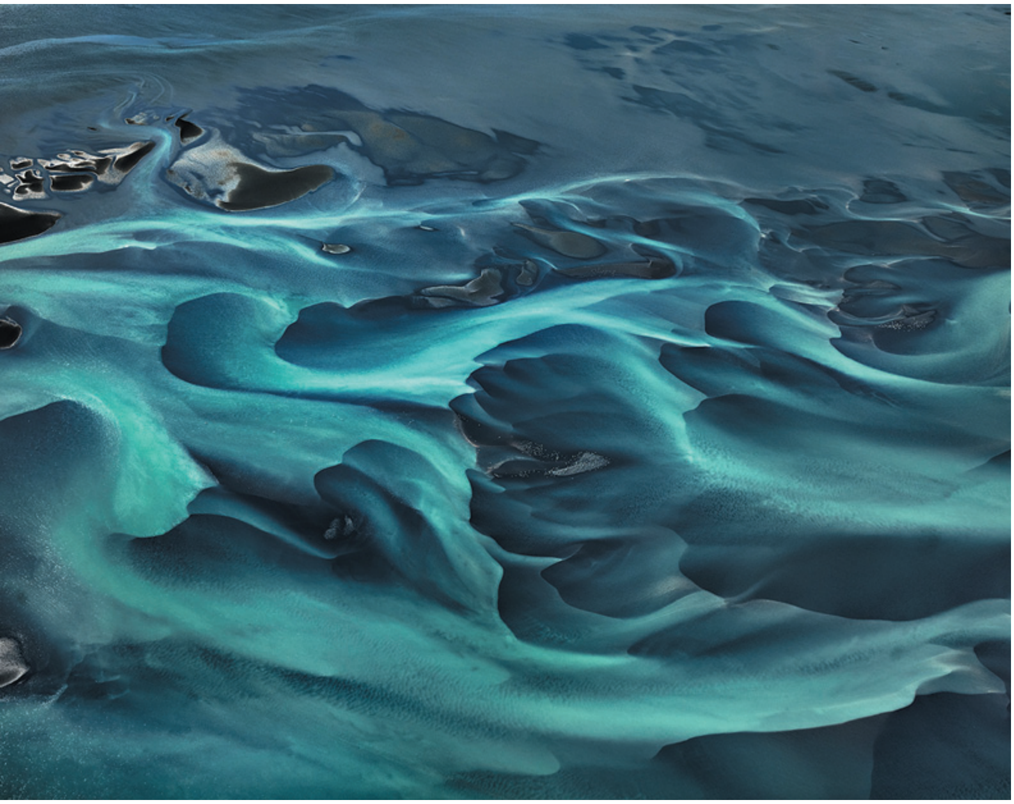
In discussing how alcoholic drinks are aged, Rogers pays particular attention to the role of barrels in producing whisky, bourbon and wine. Slow but steady exposure to low levels of oxygen, and the leaching from the oak of flavour compounds such as lactones and vanillin — which give notes of coconut and vanilla, respectively — are both important in shaping the flavour of these drinks.

Proof is an entertaining, well researched piece of popular-science writing. Rogers scrutinizes the role of smell and taste in wine tasting; this has proved a fertile ground for scientists interested in sensory perception, partly because of wine professionals' highly detailed descriptions of how they perceive taste. There is some interesting discussion about the connection between flavour perception, the words we use to describe the flavours and the chemical composition of the drink. Sensory scientists use statistical techniques to link aromas perceived by tasters with chemical analysis of wines, but it can be a daunting task. Rogers also dips into the downside: how alcohol affects our physiology, and hangovers — perhaps caused in part by the accumulation of the breakdown product of alcohol, the fairly toxic molecule acetaldehyde.

That downside is amply explored by Dudley in *The Drunken Monkey*. He, too, gives plenty of background on the history of drinking, the effects of alcohol on health and alcoholism. But the thrust of the book is an attempt to explore human alcohol use through the lens of evolution. He puts the roots of overindulgence in ancient primate tendencies to seek out ripe, sugar-rich fruits, which would often have some alcohol content because of yeast activity. Dudley, an evolutionary biologist, hypothesizes that alcohol activates neural pathways that were once nutritionally useful, but now falsely signal reward after excess consumption. This is a well constructed, clearly written book, but the overall impression is that Dudley's hypothesis, interesting as it is, needs more data points. Information on the blood alcohol levels of animals that have fed on fermenting fruit would be particularly welcome.

It is remarkable that, where not disallowed by religious beliefs, alcohol consumption has remained prevalent in so many human cultures, as both these books show. And what of the future? Most alcoholic drinks are still concocted using ancient techniques. Decades hence, will burgeoning scientific knowledge lead to new methodologies and novel forms of booze? Or will our descendants still be quaffing versions of the same old elixirs, magicked out of grain and grape by the time-honoured processes of fermentation and distillation? ■

Jamie Goode *is the author of* The Science of Wine. *He blogs at www.wineanorak.com and is based in London.*
e-mail: drjamiegoode@gmail.com



EDWARD BURTYNSKY

The Ölfusá is Iceland's largest river by volume.

DEVELOPMENT

Dammed dreams

Monya Baker is swept along by a documentary film tracing humanity's complex relationship with water.

Environmental degradation has never been so gorgeous. Over its 92 minutes, the documentary film *Watermark* deluges viewers with astonishing images of dry riverbeds, bizarre irrigation schemes, dams and city-scale aquaculture. The work of photographer and co-director Edward Burtynsky, co-director Jennifer Baichwal and cinematographer Nick de Pencier, the film is the fruit of a five-year odyssey in which Burtynsky recorded how water has shaped humanity, and vice versa. Shooting in ultra-high-definition video, he and his team specialized in aerial shots. They hoisted cameras up flagpole-high winches, and shot through the open floor of a helicopter and from airborne drones with cameras flying on a gyro-stabilized mount.

The documentary covers some 20 stories across 10 countries, filming everywhere from within a glacier to far above Earth. It shows how water is studied, diverted and polluted for profit and for pleasure, with eye-popping effects on communities and landscapes. This is no call to action, however: Burtynsky says that his images could fit as well on the cover of a mining company's corporate report as in an environmental fund-raising campaign. But the images do carry a message.

Watermark
DIRECTED BY EDWARD BURTYNSKY AND JENNIFER BAICHWAL
2013; in cinemas now.

Most start as mysteries. The film begins wordlessly, with storeys-high sepia splashes that finally resolve into a desilting project at the Xiaolangdi Dam on China's Yellow River. This cuts quickly to acres of baked, cracked mud: the ghost of a Mexican river that no longer meets the sea. "Once, the river was beautiful," a small, wrinkled woman, a former local resident, says in Spanish. As if afraid that she may be remembering a dream, Inocencia González Sainz recounts how bountiful fish were in the Colorado River Delta, now a desert because of dams — including the Hoover Dam, less than an hour's drive from Las Vegas, Nevada.

A shot of what appears to be the night sky comes slowly into focus, gradually revealed to be stacks of gleaming metal cylinders — hand-annotated in black and filled with ice cores drilled from deep within a Greenland glacier. They are also core to the film's scientific content. Jørgen Peder Steffensen and Dorthe Dahl-Jensen, who study the cores, explain what they reveal about ancient climate, and how easily climate can flip between stable states, some inhospitable to humans. Infusing water with awe, they tell how Earth's oceans were probably delivered by comets.

Burtynsky captures large-scale water-control efforts, both ancient and modern. In the steeply terraced rice paddies of China's rural Yunnan province, a teenager sporting a pink-sequined bowler hat spends his days in solitary walks as a "water guard", making sure that no one shifts the carved logs that allocate streamflow to his and other families' fields. Time-lapse photography of the construction of the Xiluodu Dam in Yunnan shows a month's worth of water turning land into reservoir; the footage then slows, zooming in on a spider that has futilely climbed to the top of an island of debris, its searching feet finding water on all sides.

Burtynsky's lens does not neglect the developed world. One sequence starts inside a chlorinated swimming pool. As the shot pans up we see that the pool is a huge tub of concrete that is itself sitting in water on a reconstructed "waterfront" in California. The curlicued streets are dotted with identical houses, many with a pool jutting into a river delta. The excess is at once disgusting and beautiful.

In a sequence on farming, a helicopter flies over a disturbingly artificial landscape of green and brown circles in Texas. They are created by long, rotating irrigation pipes that drain the Ogallala Aquifer beneath faster than it can be refilled. A helicopter pilot tells Burtynsky that a volume greater than Lake Erie has already been consumed.

Water can be trashed as well as taken. At a thirsty tannery in Dhaka, preternaturally blue waste water pours untreated into the Buriganga River. A chemist at the plant dispassionately describes the toxin-laced tanning process. In one scene, sari-clad women wearing black rubber gloves pack up piles of hide; they stand on the chromium-laced scraps in bare feet.

In a metaphysical but scientifically accurate discourse, an indigenous Canadian explains water as the stuff of spiritual connection. From a canoe in the middle of a boreal lake in British Columbia, he reminds us of a unique web of kinship in the Stikine River valley. Every living thing in it is composed mainly of a shared substance — water. ■

Monya Baker *is acting Comment editor for* Nature *in San Francisco, California.*

Correspondence

Tamiflu reviewers respond to critics

As authors of the Cochrane review that questions the stockpiling of the antiviral drugs Tamiflu (oseltamivir) and Relenza (zanamivir) against influenza pandemics, we wish to clarify aspects of your report on criticisms of the review (see *Nature* **508,** 439–440; 2014).

We agree that the randomized clinical trials of Tamiflu were "not designed to test for the severe outcomes". But far from undermining our review, this is actually one of our important findings. This is because, for years, governmental bodies justified stockpiling Tamiflu (see go.nature.com/ucyjwb and go.nature.com/oi9zbg) on the basis of a short analysis of ten pooled randomized trials (L. Kaiser *et al. Arch. Intern. Med.* **163,** 1667–1672; 2003).That study was authored by researchers at Roche, the manufacturer of Tamiflu, and concludes that the drug significantly reduces complications and hospitalizations in healthy and at-risk adults.

Our Cochrane review, by contrast, independently evaluated data from the full, previously confidential, trial-evidence base — something that officials should have done themselves. Critics of our research miss the point about what our findings say about government accountability.

You incorrectly refer to the randomized trials as "small", which would call the generalizability of the conclusions into question. In fact, trial M76001 had more than 1,400 participants, and the two pivotal studies (WV15670 and WV15671) each had more than 600 participants. You also omit to mention that the trials enrolled at-risk as well as healthy subjects.

Your report cites an observational study in which neuraminidase inhibitors (the drug class to which Tamiflu and Relenza belong) reduced mortality in hospitalized patients during the H1N1 influenza outbreak in 2009–10, apparently aligning with criticisms of our review for not including such observational studies. However, you omitted to mention the limitations of that study — or that it was funded by Roche.

We stand by our conclusion that government decisions to stockpile Tamiflu should be backed by high-quality evidence of safety and effectiveness.

Peter Doshi *University of Maryland School of Pharmacy, Baltimore, Maryland, USA.*
pdoshi@rx.umaryland.edu
Tom Jefferson *The Cochrane Collaboration, Rome, Italy.*
The authors declare competing financial interests: see go.nature.com/wudyco for details.

Prion identity wrongly credited

The review of Stanley Prusiner's autobiography (G. Mallucci *Nature* **508,** 180–181; 2014) suggests that the idea of an infectious protein was first put forward by Tikvah Alper and colleagues (*Nature* **214,** 764–766; 1967) and by John Stanley Griffith (*Nature* **215,** 1043–1044; 1967). This perpetuates a common myth.

Alper concluded from radiation-inactivation data that the agent that causes scrapie, a neurodegenerative sheep disease, does not depend on either a nucleic acid or a protein to replicate, favouring an earlier suggestion that it might be a replicating polysaccharide.

Griffith opens his paper by crediting the idea that the scrapie agent is a protein to an earlier paper by Alper and colleagues (T. Alper *et al. Biophys. Biochem. Res. Commun.* **22,** 278–284; 1966), and also to I. H. Pattison and K. M. Jones (*Vet. Rec.* **80,** 2–9; 1967). In fact, this earlier Alper paper does not contain the word 'protein'. Griffith's second claim is correct. Pattison and Jones made their suggestion because the techniques they used to purify the scrapie agent were the same as those used to purify small basic proteins.

This myth probably persists because the key 1967 papers are not freely accessible online, making it harder for today's busy scientists to check their facts.

R. John Ellis *University of Warwick, Coventry, UK.*
r.j.ellis@warwick.ac.uk

Call for UN to act on food security

The latest report from the Intergovernmental Panel on Climate Change (IPCC) indicates that the rise in greenhouse-gas emissions is affecting food production, particularly in poor tropical regions (see go.nature.com/afvyfg). As director of the CGIAR Research Program on Climate Change, Agriculture and Food Security, I call for next month's session of the United Nations Framework Convention on Climate Change to act urgently on these findings (see go.nature.com/lrwfnw). Climate-change adaptation must become the priority for policy-makers around the world.

The UN Food and Agricultural Organization (FAO) has confirmed that food prices are rising sharply (see go.nature.com/yavdzo). Recent geopolitical tensions, such as the ongoing situation in Ukraine, are partly to blame, but unusual adverse weather conditions are a main culprit.

Extreme climate events such as floods, tornadoes and droughts are becoming more frequent. Yields of wheat and maize (corn) are falling; warming oceans are wreaking havoc on fish harvests; and rising sea levels threaten to wash away fertile coastal regions. As the FAO report shows, these factors are increasing global food insecurity.

Governments have been too slow to react. Research and innovation should start now because it can take up to 20 years to see results. The UN must stop procrastinating on adaptation funding, and use the IPCC and FAO reports as an impetus for action against fractured food production (see also T. MacMillan and T. G. Benton *Nature* **509,** 25–27; 2014).

Bruce Campbell *CGIAR, Copenhagen, Denmark.*
b.campbell@cgiar.org

Ocean pollution foils search for plane

An international search to locate missing Malaysian Airlines flight MH370, which disappeared on 8 March, is under way in the southern Indian Ocean. Various objects seen floating in the ocean and washed up on the shores of western Australia briefly raised hopes that traces of the plane had been found. Unfortunately, such litter is ubiquitous in the oceans.

Finding traces of humanity in the sea has never been easier, thanks to sophisticated technology. But finding evidence of the plane's whereabouts is proving much more difficult — despite numerous and ongoing research successes with the marine-observation systems employed (see www.ioc-goos.org).

Perhaps the tragedy of the false litter trail of flight MH370 will help to raise the public's awareness of the need to protect the oceans from pollution (see www.gpa.unep.org).

Keith Alverson *United Nations Environment Program, Nairobi, Kenya.*
keith.alverson@unep.org

CONTRIBUTIONS

Correspondence may be submitted to **correspondence@nature.com** after consulting the author guidelines at **go.nature.com/cmchno**.

CLIMATE SCIENCE

Shifting storms

An analysis of historical storm data reveals that the average latitude at which tropical cyclones attain their maximum intensity has undergone a pronounced shift towards the poles over the past three decades. SEE LETTER P.349

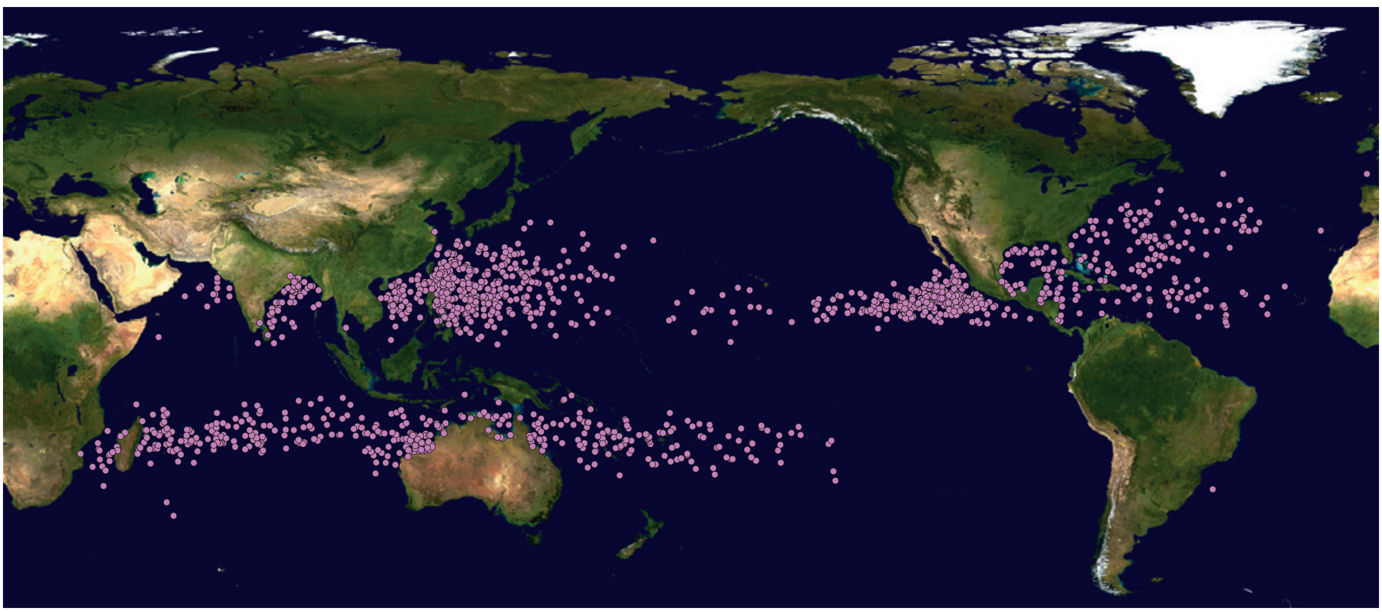
HAMISH RAMSAY

Considerable attention has been devoted to the regional and global effects of climate variability and climate change on the behaviour of tropical cyclones over the past decade or so. Catastrophic events such as Hurricane Katrina (2005), Cyclone Nargis (2008), Hurricane Sandy (2012) and Typhoon Haiyan (2013) have led scientists and non-scientists alike to ask how climate change is affecting the intensity, frequency and location of tropical cyclones around the globe. There is a general consensus among experts that anthropogenic warming will lead to fewer, but more intense, tropical cyclones[1]. However, little attention has been paid to understanding long-term shifts in the geographical location of these cyclones, particularly when at their peak intensities (Fig. 1).

On page 349 of this issue, Kossin and co-authors[2] shed light on this aspect by examining trends in the latitude at which the maximum intensities of storms occur, a metric referred to in their study as the lifetime-maximum intensity (LMI). Their findings reveal a pronounced migration of the annual-mean LMI towards the poles over the past 30 years, at a rate of about 1° of latitude per decade, although this metric varies considerably on regional scales. If this poleward migration of tropical-cyclone LMI continues, it will probably have major impacts, including increased threats to coastal communities that have historically not been susceptible to hazards posed by tropical cyclones.

The observed poleward trends in the annual-mean LMI are consistent with, and within the range of, the observed expansion of the tropics since about 1979 (refs 3, 4). Several climate-related features have been used to diagnose this expansion, which is thought to be due to increased concentrations of anthropogenic greenhouse gases. These features include ozone depletion in the stratosphere, which lies just above the lowest portion of the atmosphere (the troposphere); the height of the boundary between the stratosphere and the troposphere (the tropopause); and the width of the Hadley circulation, the main meridional overturning circulation in the troposphere, which is characterized by rising air and thunderstorms near the Equator and dry, sinking air at around 30° north and 30° south, where many of the world's deserts are found.

Kossin *et al.* suggest that two factors known to modulate tropical-cyclone development and intensity may have contributed to the observed poleward migration of annual-mean LMI: deep-layer vertical wind shear, that is, the absolute difference between wind speeds in the upper and lower troposphere; and potential intensity, a thermodynamically based theoretical upper limit of tropical-cyclone intensity that depends on local sea surface temperature and atmospheric temperature and humidity. Many storms never achieve their potential intensity because of competing influences, such as strong vertical wind shear and intrusions of dry air. However, in principle, increased potential intensity and decreased vertical wind shear should promote more-intense storms, all other factors being equal. That such trends moving away from the Equator have been observed over the past 30 years (see Fig. 2 of the paper[2])



BACKGROUND IMAGE: NASA GODDARD SPACE FLIGHT CENTER

Figure 1 | Global distribution of tropical cyclones at their peak intensities. The background image is from NASA's Visible Earth catalogue, and the tropical-cyclone data come from the National Climatic Data Center's IBTrACS archive[10,11] for the period 1982–2012. Only the locations of storms that achieved an intensity of at least a category 1 hurricane (that is, a wind speed of at least 119 kilometres per hour) are shown. The locations represent a subset of the 'best-track' data used by Kossin and colleagues[2] to construct global and regional trends in the mean latitude at which storms reached their maximum intensities.

therefore seems at least qualitatively consistent with the observed poleward migration of annual-mean LMI.

Despite the large and statistically significant global trends in the annual-mean latitude of LMI, substantial region-to-region and year-to-year variability is evident. For instance, the North Atlantic region, which has received considerable media attention owing to events such as hurricanes Katrina and Sandy, shows almost no poleward trend on the basis of historical 'best-track' data over the past 30 years. Moreover, when the authors used a state-of-the-art data set of tropical-cyclone intensity (ADT-HURSAT; ref. 5), an opposite, equatorward, trend is found for the North Atlantic (see Table 1 of the paper[2]). Such regional differences in trends are probably due to climate modes that extend in time beyond the period for which accurate satellite-based data are available.

This is one of the limitations of trend studies based on satellite-derived estimates of tropical-cyclone intensity. Although the post-1970s geostationary satellite era is considered to be the most accurate part of the historical tropical-cyclone record, the relatively short observation period hampers the detection of trends influenced by modes of climate variability whose periodicity spans decades or longer, such as the Pacific Decadal Oscillation[6]. Any such variability implies that regions in which the poleward migration of annual-mean LMI has been more pronounced over the past 30 years might experience less-pronounced trends in the coming decades, and vice versa. Even on a global scale, a trend of 1° of latitude per decade of tropical expansion (that is, a 10° shift per century, assuming a constant rate of expansion) cannot be sustained without implausible changes to fundamental physical constraints on the global atmospheric circulation, such as Earth's rotation rate.

On year-to-year timescales, variability in tropical-cyclone formation and track is dominated by the phase of the El Niño–Southern Oscillation (ENSO) — the episodic warming (El Niño) and cooling (La Niña) of the surface temperature of the tropical Pacific Ocean. El Niño often promotes an equatorward migration of tropical-cyclone activity, whereas during La Niña a poleward displacement is observed[7], concomitant with changes in the width and intensity of the Hadley circulation[8]. It is therefore plausible that any trend in ENSO could project onto trends in tropical-cyclone activity. Kossin *et al.* attempt to remove this contribution by accounting for the effect of ENSO on the linear trend of annual-mean LMI latitude and then examining the residual data. The poleward migration remains pronounced and statistically significant, suggesting that ENSO plays only a minor part in the long-term hemispheric and global trends.

Kossin and colleagues' findings provide insight into the response of global tropical-cyclone activity to a changing climate. However, several questions remain unanswered. For instance, will future changes in wind patterns cause storms to move towards or away from coastlines[9]? What are the key mechanisms driving the observed tropical expansion, and how do these tie in with factors known to modulate tropical-cyclone intensity? Such questions remain the subject of future research. ■

Hamish Ramsay *is at the ARC Centre of Excellence for Climate System Science and the School of Mathematical Sciences, Monash University, Victoria 3800, Australia.*
e-mail: hamish.ramsay@monash.edu

1. Knutson, T. R. *et al. Nature Geosci.* **3,** 157–163 (2010).
2. Kossin, J. P., Emanuel, K. A. & Vecchi, G. A. *Nature* **509,** 349–352 (2014).
3. Hartmann, D. L. *et al.* in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Stocker, T. F. *et al.*) Ch. 2, 226–229 (Cambridge Univ. Press, 2013).
4. Seidel, D. J., Fu, Q., Randel, W. J. & Reichler, T. J. *Nature Geosci.* **1,** 21–24 (2008).
5. Kossin, J. P., Olander, T. L. & Knapp, K. R. *J. Clim.* **26,** 9960–9976 (2013).
6. Grassi, B., Redaelli. G., Canziani, P. O. & Visconti, G. *J. Clim.* **25,** 3282–3290 (2012).
7. Ramsay, H. A., Camargo, S. J. & Kim, D. *Clim. Dyn.* **39,** 897–917 (2012).
8. Nguyen, H., Evans, A., Lucas, C., Smith, I. & Timbal, B. *J. Clim.* **26,** 3357–3376 (2013).
9. Barnes, E. A., Polvani, L. M. & Sobel, A. H. *Proc. Natl Acad. Sci. USA* **110,** 15211–15215 (2013).
10. www.ncdc.noaa.gov/ibtracs
11. Knapp, K. R., Kruk, M. C., Levinson, D. H., Diamond, H. J. & Neumann, C. J. *Bull. Am. Meteor. Soc.* **91,** 363–376 (2010).

SYNTHETIC BIOLOGY

New letters for life's alphabet

The five bases found in nucleic acids define the 'alphabet' used to encode life on Earth. The construction of an organism that stably propagates an unnatural DNA base pair redefines this fundamental feature of life. SEE LETTER P.385

ROSS THYER & JARED ELLEFSON

All known life forms store and transmit information from generation to generation using the bases found in nucleic acids: adenine, cytosine, guanine, thymine and uracil. In nucleic-acid double helices, these form base pairs (guanine with cytosine, and either adenine with thymine in DNA, or adenine with uracil in RNA), which are mostly orthogonal — that is, little pairing occurs between other combinations of bases. However, this 'alphabet' seems to be an accident of history rather than a functional necessity, given that other orthogonal base pairs have been synthesized and shown to be processed by DNA-replication enzymes *in vitro*[1]. Because life on Earth is biochemically uniform, the formal possibility of alternative alphabets requires strong experimental proof. In this issue, Malyshev *et al.*[2] (page 385) provide just such a proof, by conclusively showing that an unnatural base pair can be stably propagated in the bacterium *Escherichia coli*.

Shortly after the discovery of DNA, it was proposed[3] that analogues of natural bases could form a third functional pair, but nearly 30 years passed before advances in organic synthesis and the development of methods for amplifying DNA gave scientists free reign to explore this hypothesis. In 1989, a base pair formed from isomers of guanine and cytosine was synthesized, and replication, transcription and even translation of DNA sequences incorporating this base pair were demonstrated *in vitro*[1,4]. Then in 1995 came the surprising finding[5] that hydrogen bonding between bases was not an absolute requirement for complementary binding, and could be replaced by steric compatibility (the fitting together of matching molecular shapes) and hydrophobic interactions. This culminated in the independent development of three highly orthogonal base pairs[6–8], each capable of *in vitro* replication fidelity exceeding 99%.

Malyshev *et al.* now describe the development of a bacterium capable of faithfully replicating a plasmid — a small, circular DNA molecule — containing the hydrophobic d5SICS:dNaM base pair (Fig. 1), thus creating the first organism to harbour an engineered and expanded genetic alphabet. This feat was far from simple: the authors first had to find a way of getting the bacterium to take up unnatural nucleotides, and then to work within the constraints of the billion-year-old habits of polymerases, the enzymes that synthesize polymeric nucleic acids.

➲ NATURE.COM
For more on the expanded genetic alphabet, visit:
go.nature.com/gmcbeg

To solve the first problem, Malyshev and colleagues engineered an *E. coli* strain that expressed an algal nucleotide triphosphate transporter

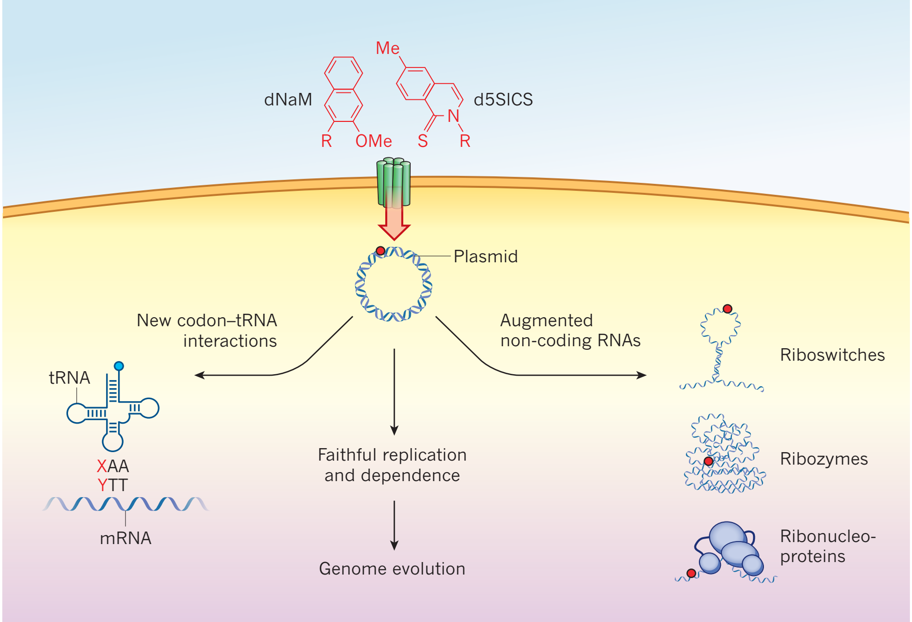


Figure 1 | Prospects for organisms that propagate unnatural DNA base pairs. Malyshev *et al.*[2] have generated a strain of the bacterium *Escherichia coli* that expresses an algal transporter protein (green), which imports the unnatural nucleotide triphosphates dNaM and d5SICS from the culture medium. This allows the bacteria to replicate a plasmid that incorporates the unnatural dNaM:d5SICS base pair (red dot). In turn, this might open up many further developments, including: organisms that can add new codons to the genetic code through customized codon–transfer-RNA interactions (XAA and YTT are the anticodon and codon of the tRNA and messenger RNA, respectively; X and Y are unnatural bases, A and T are natural bases); organisms that faithfully replicate and depend on unnatural base pairs, which might allow genome evolution; and non-coding RNAs (such as riboswitches, ribozymes and those in ribonucleoproteins) that have augmented functions. Me, methyl group; R represents the sugar and phosphate groups of the nucleotide triphosphates.

(NTT) protein, which allowed direct import of the nucleotides d5SICS and dNaM. To ensure efficient replication, the authors placed the unnatural base pair in a region of a plasmid predicted to be replicated solely by DNA polymerase I. Rather than being the workhorse of DNA replication, this enzyme fills in gaps in DNA molecules or connects 'Okazaki' DNA fragments, and has been shown[9] to replicate the d5SICS:dNaM pair efficiently *in vitro*.

After introducing a plasmid containing a single d5SICS:dNaM base pair into *E. coli* and supplementing the media with the two unnatural nucleotides, the researchers demonstrated that the unnatural base pair was retained in the plasmid after days in culture. They proved the presence of the unnatural base pair in recovered plasmids using a battery of techniques. Retention of the unnatural base pair after 15 hours of cell growth and plasmid replication was estimated to be at least 99.4% per doubling of the plasmid, an error rate no worse than that of some viral polymerases.

The next step will be to ensure long-term retention, which may require the engineering of a bacterium that depends on the unnatural base pair. It may be that the biological machinery used in Malyshev and colleagues' *E. coli* will allow the organism to readily adopt the unnatural bases as part of its own genetic alphabet. If so, this would open up a new vista in which human engineering can leap chasms previously unfathomable to evolution. This may seem fanciful, but wholescale reassignment of the genetic code produced through several billion years of evolution also seemed unlikely, and has nonetheless recently been achieved[10].

Once unnatural base pairs are not just tolerated by an organism, but also accepted and used, the next crucial step will be to demonstrate that they can be transcribed into RNA *in vivo*. From there, the opportunities multiply quickly (Fig. 1) — for example, unnatural nucleotide pairs might augment functional RNA elements, such as riboswitches and ribozymes. The incorporation of unnatural nucleotides into DNA promoter sequences or repressor binding sites (which initiate or subdue gene expression, respectively, by acting as binding sites for proteins), in conjunction with engineering of their partner proteins, might be used to formulate new and independent regulatory architectures.

Similar engineering feats could also provide unique functionality to RNA–protein complexes, for instance, by restricting the binding of the Cas9 enzyme (a widely used tool for generating double-strand breaks in DNA) to sites containing an unnatural base pair. But perhaps the ultimate application of such base pairs will be to add novel codons — triplets of nucleotides that encode which amino acids are incorporated into proteins — to the genetic code through customized codon–transfer-RNA interactions. This would greatly expand the number of available codons that can be assigned new translational functions, such as encoding non-standard amino acids, and would prevent synthetic biologists from having to recode the translational functions of existing codons[10] through painstaking genome engineering. In other words, an expanded genetic alphabet will help build an expanded translational alphabet.

But why stop at six letters in DNA? The NTT used by Malyshev *et al.* may be fairly promiscuous, importing both natural and unnatural nucleotides indiscriminately. Other groups have developed unnatural base pairs[7,8] that could be equally acceptable substrates for the transporter and for the cellular replication machinery. If the technique for introducing d5SICS:dNaM into *E. coli* works for other pairs, then the DNA code could be extended well beyond three base pairs. This raises fundamental questions about why life settled on only two in the first place, and whether semi-synthetic organisms with the capacity to store more information will have expanded capabilities (as we envisage above) or endure intolerable fitness costs (owing to inherently lower fidelity of DNA replication, RNA misfolding or translation-error catastrophes).

Attempts to expand the genetic alphabet bravely question the idea of the universal nature of DNA, and potentially draw criticism about the wisdom of tinkering with it. Such criticisms should be solidly countered by synthetic biologists at the outset. James Watson and Francis Crick's discovery of base pairing in DNA yielded a mechanism for genetics, but now genetics has inexorably yielded a mechanism for greater biological diversity, and thus potentially for building a better biological future. ■

Ross Thyer *and* **Jared Ellefson** *are in the Center for Systems and Synthetic Biology, University of Texas at Austin, Austin, Texas 78712-1095, USA.*
e-mail: ross.thyer@utexas.edu

1. Switzer, C., Moroney, S. E. & Benner, S. A. *J. Am. Chem. Soc.* **111,** 8322–8323 (1989).
2. Malyshev, D. A. *et al. Nature* **509,** 385–388 (2014).
3. Rich, A. in *Horizons in Biochemistry* (eds Kasha, M. & Pullman, B.) 103–126 (Academic, 1962).
4. Bain, J. D., Switzer, C., Chamberlin, A. R. & Benner, S. A. *Nature* **356,** 537–539 (1992).
5. Schweitzer, B. A. & Kool, E. T. *J. Am. Chem. Soc.* **117,** 1863–1872 (1995).
6. Malyshev, D. A. *et al. Proc. Natl Acad. Sci. USA* **109,** 12005–12010 (2012).
7. Yamashige, R. *et al. Nucleic Acids Res.* **40,** 2793–2806 (2012).
8. Yang, Z., Chen, F., Alvarado, J. B. & Benner, S. A. *J. Am. Chem. Soc.* **133,** 15105–15112 (2011).
9. Seo, Y. J., Hwang, G. T., Ordoukhanian, P. & Romesberg, F. E. *J. Am. Chem. Soc.* **131,** 3246–3252 (2009).
10. Lajoie, M. J. *et al. Science* **342,** 357–360 (2013).

This article was published online on 7 May 2014.

ORGANIC CHEMISTRY

Collaborative synthesis

A chemical synthesis has led to the reassignment of the molecular structure of the naturally occurring compound citrinalin B. This, in turn, helps to untangle the biochemical origins of an intriguing family of natural products. See Article p.318

JOHN L. WOOD

Synthetic organic chemistry has the potential to provide convenient access to any complex molecule with atomic precision. Although this promise is still far from being fully realized, the progress made since Friedrich Wöhler's preparation of urea — the first synthesis of a naturally occurring molecule — nearly two centuries ago has been astonishing, and has enabled tremendous advances in agricultural, medicinal and materials chemistry. In this issue, Mercado-Marin *et al.*[1] report the synthesis of the natural products citrinalin B and cyclopiamine B. Their work beautifully illustrates how natural products continue to inspire research in synthesis, and shows that these efforts can benefit greatly from collaboration with experts in computational chemistry and biosynthesis.

This particular collaboration, led by the organic chemist Richmond Sarpong, was inspired by cyclopiamines A and B and citrinalins A and B, four closely related members in the growing class of prenylated indoles[2,3]. Compounds in this class are generally derived from fungi and have a wide range of intriguing biological properties, so their synthesis and biosynthesis have been the focus of considerable research. These efforts have established that compounds similar to cyclopiamine A and citrinalin A probably derive from a common biosynthetic precursor that differentiates into enantiomers (non-superimposable mirror-image isomers) of an intermediate compound, each of which serves as a precursor to a particular product. This notion is supported by the fact that large portions of the structures of cyclopiamine A and citrinalin A are mirror images of each other (Fig. 1); the complete mirror-image symmetry of the enantiomeric precursors is destroyed in the course of the natural products' biogenesis, and so cyclopiamine A and citrinalin A are said to be only pseudo-enantiomeric[4,5].

Cyclopiamine B is thought to derive from cyclopiamine A, and Sarpong and colleagues speculated that citrinalin B similarly derives from citrinalin A. This would result in a pseudo-enantiomeric relationship between citrinalin B and cyclopiamine B. However, when they compared the reported structures of cyclopiamine B and citrinalin B, confusion arose because the core structures are diastereomers (isomers that differ in the three-dimensional orientations of their atoms, but that are not mirror images). Although diastereomeric relationships are not uncommon among natural products, it was difficult for the researchers to envisage a logical mechanism whereby citrinalin A could give rise to citrinalin B. Unable to reconcile the biosynthetic origin of citrinalin B, Sarpong and colleagues questioned its reported structure and postulated a revised structure through which the pseudo-enantiomeric relationship between cyclopiamine B and citrinalin B is restored (Fig. 1).

To provide support for this hypothesis, Sarpong consulted with the computational chemist Dean Tantillo, who, along with a colleague, calculated thermodynamic properties and simulated spectra for the cyclopiamines and citrinalins (original and revised structures). The simulated data for the revised structure of citrinalin B best matched those obtained experimentally for the isolated natural material. Moreover, the calculations predicted that the thermodynamically preferred products are, by a large margin, cyclopiamine B and citrinalin B. Thus, if the B isomers form biosynthetically from the corresponding A isomers, one would expect to isolate little, if any, of the A isomers. The relatively high abundance of the A isomers isolated from natural sources led Sarpong and co-workers to speculate that cyclopiamine B and citrinalin B may be artefacts of the conditions used in the isolation process.

Having obtained computational support for the hypothesized structural revision, Sarpong and colleagues devised a synthesis of 'revised' citrinalin B to provide unambiguous confirmation of its structure. They used the amino acid D-proline as the point of departure in a 16-step sequence that delivered an indole intermediate (Fig. 2). In a key step of the synthesis, the researchers then had to convert this intermediate to the corresponding oxindole, a compound that contains a synthetically challenging carbon atom carrying four different groups, known as a stereogenic quaternary carbon. To ensure the conversion, the authors postulated that the indole would need first to undergo a diastereoface-selective epoxidation (selective introduction of an oxygen to the 'back' face of the molecule; the resulting epoxide group is shown in blue in Fig. 2). This would produce an intermediate that is poised for regioselective ring opening (selective breaking of the carbon–oxygen bond in the epoxide that is distal to the adjacent indole nitrogen) and subsequent migration of a carbon–carbon bond, controlled so as to set up a specific three-dimensional arrangement of atoms at the stereogenic carbon.

The indole-to-oxindole transformation was fruitful, but its development required considerable experimentation, and success relied on several key structural features present in the indole substrate. Specifically, a primary amine group (NH_2) was crucial in guiding the facial selectivity observed during the initial epoxidation step, and a chromanone unit (highlighted

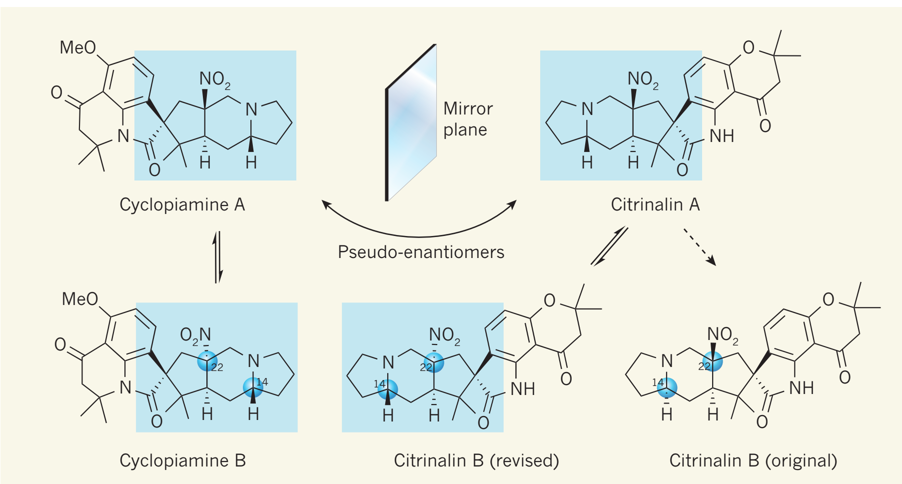


Figure 1 | A structural inconsistency. Cyclopiamines and citrinalins are members of the prenylated indole class of natural products. Cyclopiamine A converts to its B isomer in a reversible reaction, and is a pseudo-enantiomer of citrinalin A — that is, the boxed section of cyclopiamine A is the mirror image of the analogous section of citrinalin A. Mercado-Marin *et al.*[1] argued that citrinalin B must form from citrinalin A in the same reversible reaction, but the structure originally assigned to the B isomer is not a pseudo-enantiomer of cyclopiamine B, as would be expected (carbons 14 and 22 have the wrong orientations). The authors therefore proposed a revised structure for citrinalin B, which is a pseudo-enantiomer of cyclopiamine B. Me, methyl.

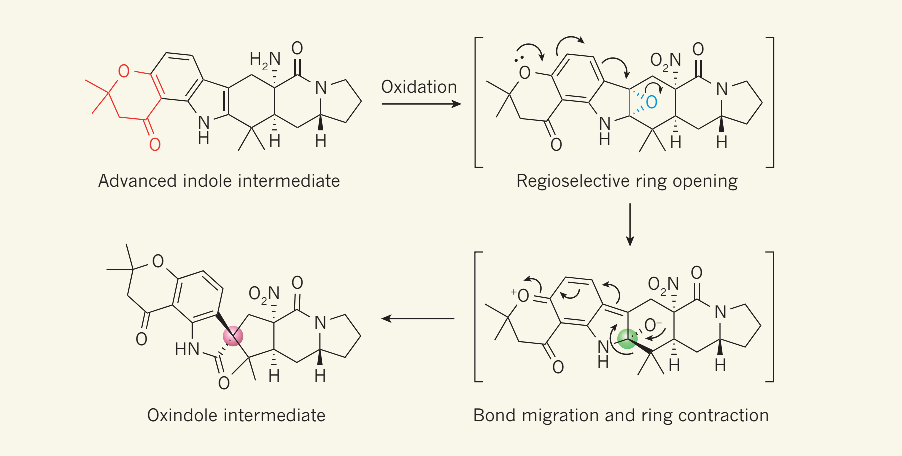


Figure 2 | Key step in the synthesis of ent-citrinalin B. In Mercado-Marin and colleagues' synthesis[1] of the revised structure of ent-citrinalin B, a key step was to convert an indole intermediate to an oxindole intermediate, ensuring the formation of the three-dimensional arrangement — the stereochemistry — of groups around the stereogenic centre, which is highlighted in pink in the oxindole. (A chromanone group in the indole is shown in red.) The authors used an oxidation reaction to attach an oxygen atom to the 'back' face of the indole, forming an epoxide ring (blue). A regioselective ring-opening reaction then occurred, facilitated by a pair of electrons (dots) from an oxygen atom in the chromanone; curly arrows indicate electron movement. This reaction set up the desired stereochemistry around the carbon atom (highlighted in green) that goes on to form the stereogenic centre of the oxindole. Subsequent bond migration and ring contraction led to the formation of the target oxindole. Square brackets indicate transiently formed reaction intermediates.

in red in Fig. 2) facilitated regioselective opening of the presumed intermediate epoxide[6]. In the latter event, an electron pair on the chromanone might help to break the distal carbon–oxygen epoxide bond, whereas undesirable competition by the indole nitrogen's lone pair is mitigated by the chromanone's carbonyl group (C = O), which withdraws electron density from the nitrogen.

Sarpong and colleagues readily advanced the resulting oxindole intermediate to make the enantiomer of citrinalin B (ent-citrinalin B), which, in turn, provided access to cyclopiamine B through a rearrangement of the chromanone group. Importantly, the spectral data for these compounds were identical to those of naturally occurring samples (with the exception of the optical rotation of ent-citrinalin B, the sign of which was opposite to that for natural citrinalin B, as expected for the enantiomer), thus confirming the postulated structural revision.

Although most synthetic chemists would have stopped at this stage, Sarpong chose to take things further. In collaboration with the group of Roberto Berlinck, the chemist who originally isolated the citrinalins, Sarpong initiated a series of studies to determine the biogenetic precursors of citrinalins and cyclopiamines. This proved to be highly successful, not only establishing the major sources of carbon atoms, but also resulting in the isolation of citrinalin C — the structure of which supports not only the hypothesis of enantiomeric precursors for citrinalins and cyclopiamines, but also the postulated biosynthetic link between these compounds that initially led the authors to propose the structural revision of citrinalin B. ■

John L. Wood *is in the Department of Chemistry and Biochemistry, Baylor University, Waco, Texas 76798, USA.*
e-mail: john_l_wood@baylor.edu

1. Mercado-Marin, E. V. *et al. Nature* **509,** 318–324 (2014).
2. Bond, R. F., Boeyens, J. C. A., Holzapfel, C. W. & Steyn, P. S. *J. Chem. Soc. Perkin Trans. 1* 1751–1761 (1979).
3. Pimenta, E. F. *et al. J. Nat. Prod.* **73,** 1821–1832 (2010).
4. Porter, A. E. A. & Sammes, P. G. *J. Chem. Soc. D* 1103 (1970).
5. Finefield, J. M., Frisvad, J. C., Sherman, D. H. & Williams, R. M. *J. Nat. Prod.* **75,** 812–833 (2012).
6. Zhang, X. & Foote, C. S. *J. Am. Chem. Soc.* **115,** 8867–8868 (1993).

NEUROBIOLOGY

To care or not to care

The behaviour of adult mice towards pups varies depending on gender and sexual experience. The activity of a population of neurons in the hypothalamus of the brain has now been found to regulate these differing responses. SEE ARTICLE P.325

IVAN RODRIGUEZ

The survival of vertebrates depends on their ability to display appropriate behaviour towards nearby individuals — for example, towards predators, prey or members of their own species. In some cases, escaping death depends on being able to trigger stereotyped responses in others. The immediate survival of newborn mammals, for instance, relies on their ability to induce a caring social response in their parents. On page 325 of this issue, Wu *et al.*[1] identify a subset of neurons in the hypothalamus of the mouse brain that is essential for parental care.

For more than 100 million years, milk has been the primary source of nutrition for newborn mammals, which means that a daily physical link between mother and young is mandatory for newborn survival[2]. Thus, by not only gestating the fetus but also suckling the newborn offspring, mammalian mothers make a much larger investment in their progeny than mammalian fathers. The contribution of males to parenting is highly variable, and depends on the species and experience of the animal. In the presence of a pup, a male mouse without sexual experience might show indifference, but the norm is a physical attack on the younger animal. This all changes after mating: the experienced male's behaviour switches from aggressive to parental.

What a mouse perceives when it encounters another mouse is largely processed by a specialized sensor in its nose — the vomeronasal organ. This structure is linked to innate responses, including fight, flight and sex-specific behaviours, a property that led Wu and colleagues to propose that the vomeronasal organ might be involved in regulating the variable predisposition of male mice to parenting. To test this theory, the authors impaired vomeronasal signalling to the brain in virgin male mice, by genetically deleting an ion-channel protein (TRPC2) that is expressed only in the vomeronasal organ. The males behaved much less aggressively towards pups than their wild-type counterparts, suggesting that signalling from the vomeronasal organ promotes aggression, directly or indirectly, in virgin males.

Wu and co-workers observed that a population of neurons in the medial preoptic area (MPOA) of the hypothalamus was activated in males and females that exhibited parental behaviour towards pups. But the neurons were less active in virgin males in the presence of newborns, suggesting that this neuronal population might be involved in promoting

caring behaviours and might be modulated by vomeronasal signalling. This could have been the end of an already interesting story had the authors not found that a large portion of the neurons expressed the neuropeptide galanin. This finding enabled them to develop a tool with which to evaluate the role of these neurons in the regulation of parenting.

The experiments that followed were simple, at least conceptually, and involved two approaches. To disrupt the function of the galanin-expressing neurons in the MPOA, Wu *et al.* genetically engineered these cells to die. In a parallel approach, they engineered the neurons to express the protein Channelrhodopsin-2, which allowed their specific activation in response to light.

The results of these two complementary experiments were striking. First, depletion of MPOA galanin-expressing neurons in virgin females caused them to behave aggressively towards pups. Similarly, sexually experienced males and females showed deficits in parenting behaviour after galanin-expressing neurons were depleted. More remarkably, light stimulation of the neurons in virgin males inhibited their attacks on newborns, and even induced pup grooming (Fig. 1). These impressive data point to a crucial role for galanin-expressing neurons in mediating parental behaviour towards pups, thus linking these neurons to the survival chances of young mice.

Several neural-circuit models can be built on the authors' findings. To take the simplest, galanin-expressing neurons may act as a regulatory node that puts the brakes on a default aggressive state promoted by signalling from the vomeronasal organ. Alternatively, these neurons might simply promote parenting behaviour.

But naturally, questions arise. For example, given the known roles of various neuropeptides in modulating behaviour[3], could galanin itself be more than just a handy genetic marker for these neurons? Supporting this hypothesis, injection of galanin into the MPOA of Syrian hamsters leads to changes in scent-marking behaviour[4].

It seems that the galanin-expressing circuitry affects more than just parenting behaviour in mice. Wu and colleagues found that some social behaviours unrelated to pups, such as inter-male aggression, were also affected by activation of the neurons. This observation makes the story more complex, but it also opens up the possibility of dissecting the neural circuits that drive distinct behaviours.

The switch between infanticide and parenting behaviour is not specific to rodents — it is pervasive in the animal kingdom, from birds to marine mammals, in both males and females. The best-known example is male lions, which kill young cubs when entering a new social group (in addition to evicting adult males and sub-adults)[5]. What is the advantage of this infanticide? One obvious explanation is that, for the male to maximize the chances of producing his own offspring, it is beneficial to ensure that female lactation ends and ovulation restarts as soon as possible.

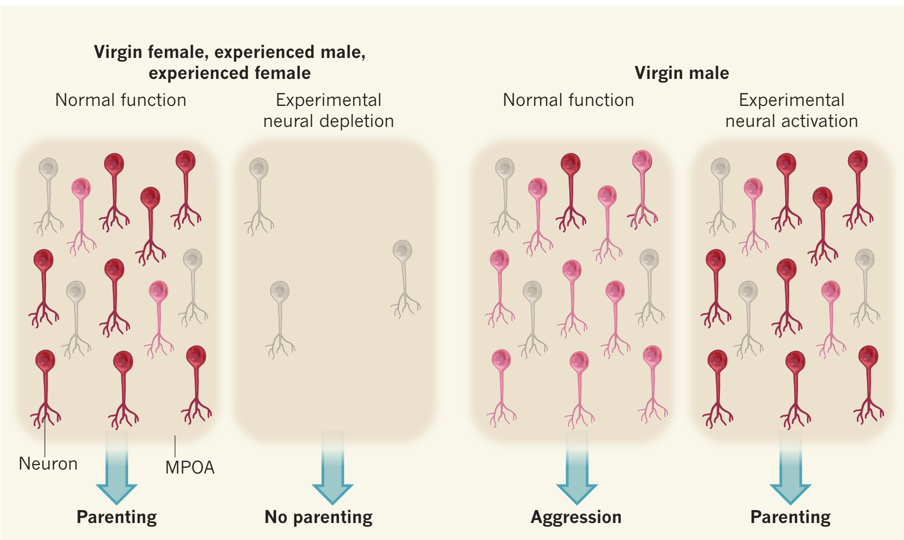


Figure 1 | Modulating parenting behaviours in the presence of pups. Virgin female mice and sexually experienced males and females show parenting behaviour towards young pups, but virgin males act aggressively towards young. Wu *et al.*[1] investigated the basis for these varying responses in the four groups of mice. They report that the responses are modulated by a group of neurons in the medial preoptic area (MPOA) of the brain that express the protein galanin (inactive galanin-expressing neurons are pink, active ones are red, and neurons that do not express galanin are grey). These neurons are present in all four groups, but in the presence of pups, fewer are active in virgin males. Depletion of galanin-expressing neurons in virgin females or in sexually experienced mice of either sex causes impaired parental responses in the presence of pups and, in some cases, aggressiveness. Conversely, activation of galanin-expressing neurons in virgin males suppresses aggression and induces pup grooming.

NATURE

50 Years Ago

'The Mathematical Association Annual Conference 1964' — Administrative problems of course loomed very large … It was easy to fail to appreciate the possible influence of the liberal and cultural qualities of mathematics on sixth formers, who were given opportunities to read round their subject. The learning of mathematics was organic and would grow wherever it was given room; like angling, it could "never be fully learnt" … It was a sombre fact that nearly 25 per cent of persons entering training colleges in 1962 did not possess a pass at the Ordinary Level of the General Certificate of Education in mathematics … The discussion from the floor represented the confusion which exists in most teachers' minds as to how one can reconcile teaching logic with geometry, when the only logical way seemed to be to start at the middle, work on establishing the standard results, and then work back and establish the premises … the discussion had shown that Euclid, instead of being found at fault, had proved himself to be too good for the age group for which we tend to use him. Unfortunately we had found nothing to replace him for this age group.

From *Nature* 16 May 1964

100 Years Ago

The annual report of the Hampstead Scientific Society … contains … a summary of the meteorological statistics for the Hampstead Observatory for 1913 … For the first time, average meteorological data are included in the report … From these preliminary averages it would appear that Hampstead is the coldest, rainiest, snowiest, and frostiest, as well as almost the sunniest and foggiest of the stations in the neighbourhood of London.

From *Nature* 14 May 1914

It is unclear whether infanticide provides such benefits to virgin male mice, given the fast weaning times and frequent oestrous cycles that are characteristic of this species. But aggressive behaviours spontaneously re-emerge in male mice 50 days after mating[6]: exactly the length of time it takes for their progeny to be born and weaned. Furthermore, exposure of a pregnant female to the scent of an unfamiliar male mouse is sufficient to cause termination of pregnancy[7]. As such, mice may be the champions of infanticide.

A picture is emerging in which regulatory nodes of social interactions switch on specific neural circuits at the expense of others[8,9]. These circuits underlie stereotyped behaviours, and coexist in both males and females, whether they are sexually experienced or not. But only under specific conditions are they activated. This is a remarkable example of the modularity and versatility of mammalian brains. ■

Ivan Rodriguez *is in the Department of Genetics and Evolution, University of Geneva, CH-1205 Geneva, Switzerland.*
e-mail: ivan.rodriguez@unige.ch

1. Wu, Z., Autry, A. E., Bergan, J. F., Watabe-Uchida, M. & Dulac, C. G. *et al. Nature* **509,** 325–330 (2014).
2. Lefèvre, C. F., Sharp, J. A. & Nicholas, K. R. *Annu. Rev. Genom. Hum. Genet.* **11,** 219–238 (2010).
3. Bridges, R. S., DiBiase, R., Loundes, D. D. & Doherty, P. C. *Science* **227,** 782–784 (1985).
4. Ferris, C. F., Delville, Y., Bonigut, S. & Miller, M. A. *Brain Res.* **832,** 1–6 (1999).
5. Pusey, A. E. & Packer, C. in *Infanticide and Parental Care* (eds Parmigiani, S. & Vom Saal, F. S.) 277–299 (Harwood, 1994).
6. Perrigo, G., Belvin, L. & Vom Saal, F. S. *Chronobiol. Int.* **9,** 421–433 (1992).
7. Bruce, H. M. *Nature* **184,** 105 (1959).
8. Kimchi, T., Xu, J. & Dulac, C. *Nature* **448,** 1009–1014 (2007).
9. Ferrero, D. M. *et al. Nature* **502,** 368–371 (2013).

SENSORY BIOLOGY

Radio waves zap the biomagnetic compass

Weak radio waves in the medium-wave band are sufficient to disrupt geomagnetic orientation in migratory birds, according to a particularly well-controlled study. But the underlying biophysics remains a puzzle. **See Letter p.353**

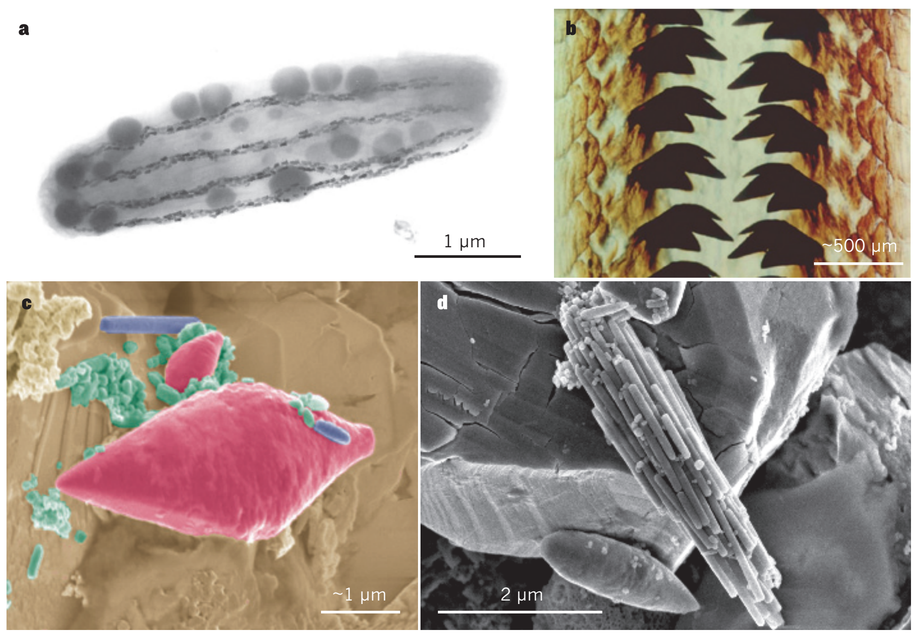
JOSEPH L. KIRSCHVINK

Magnetobiology has largely been viewed as a stamping ground for charlatans since the followers of physician Franz Anton Mesmer failed to cure patients using a 'magnetized' tree in the eighteenth century. Numerous discoveries have begun to change that perspective, although the road has been rocky. For example, early studies suggesting that migrating animals use geomagnetic cues for navigation were plagued by variability, but it is now clear that many microorganisms and animals use a magnetic compass for part of their orientation[1].

On the fringe of this fringe field were claims that radio-frequency radiation could have biological effects at levels too weak to act through the understood mechanisms of tissue heating or shock, but the experiments usually lacked proper controls and blinding techniques[2–4]. Now, however, on page 353 of this issue, Engels *et al.*[5] demonstrate convincingly that migrating European robins stop using their magnetic compasses in the presence of extraordinarily weak, radio-frequency electromagnetic 'noise'.

Using rigorous, double-blinded experiments, the authors found that birds housed in huts screened from background electromagnetic noise were able to use their magnetic compass to orient themselves appropriately, but that their orientation was disrupted following the introduction of electromagnetic noise ranging from 20 kilohertz to 5 megahertz, at intensities similar to that measured for background anthropogenic noise in the environment. To put it into perspective, this is in the medium-wave band used for AM radio transmissions (not, for example, mobile phones), and the strength is about equivalent to what a bird in flight might experience 5 kilometres away from a 50-kilowatt AM radio station.

Two results flag this study as particularly noteworthy, and puzzling. First, the levels of radio-frequency radiation that affected the birds' orientation are substantially below anything previously thought to be biophysically plausible, and far below levels recognized as affecting human health. Second, the authors detect no trace of a sharply enhanced effect at the Larmor frequency (the natural period at which single electrons wobble around the geomagnetic-field direction), which flatly contradicts experiments on the same species performed using a similar protocol[6]. This failure



A, HOJATOLLAH VALI; B, HEINZ A. LOWENSTAM/JOSEPH L. KIRSCHVINK; C,D, DIRK SCHUMANN

Figure 1 | Biological 'magnetomonsters'. Several fossil and extant organisms contain highly magnetic structures. Examples include: **a**, *Magnetobacter bavaricum*, a magnetotactic bacterium with nearly 100 times more magnetite in its cells than more typical types; **b**, *Cryptochiton stelleri*, a mollusc whose magnetite-capped radular teeth will stick strongly to a hand magnet; **c**, a spearhead-shaped magnetite particle (false-coloured red), prismatic magnetite rods (purple) and typical magnetite-containing bacterial organelles (magnetosomes; green); **d**, a bundle of magnetite rods forming 'wires'. The structures shown in **c** and **d** were extracted from fossilized clay sediments in New Jersey dating to approximately 56 million years ago[11]. The origins of the spearhead- and rod-shaped objects are not known, but their size and morphology suggest that they might have belonged to more-complex organisms. Cellular structures containing enough electrically conducting magnetite could be sensitive to radio-frequency radiation at levels shown by Engels *et al.*[5] to disrupt birds' geomagnetic orientation.

to replicate that effect perhaps underscores previously suggested[2,3] flaws in the blinding of earlier studies.

So what might be going on in these birds? Several other external stimuli that stop animals from responding to geomagnetic cues have been identified. Early studies of animal navigation noted that cues from the Sun or stars would take precedence over magnetic cues, leading to the idea that magnetism is the compass of last resort. It was then noticed that robins would ignore the magnetic field when the background intensity was shifted 20–30% outside the normal value[1], and that pigeons raced poorly during geomagnetic storms. From an evolutionary perspective, ignoring geomagnetic cues at such times makes sense, because anomalies in the background field are often associated with iron deposits or lightning strikes. Some animals also stop using their magnetic compass in the presence of red-only light, but such light is present only at sunrise and sunset, when the Sun compass is most reliable[3].

Hence, radio-frequency noise might be just another cue that tells migrating animals to ignore their magnetic sense, but the puzzle is why this might have evolved. Surprisingly, there is a natural source of the radio-frequency electromagnetic noise identified as disruptive by Engels and colleagues — that produced by solar storms. Coronal mass ejection (CME) events from the Sun slam plasma into Earth's magnetosphere every now and then, causing it to 'sing' at frequencies from as low as around 20 kHz up to the MHz range[7], some of which even leaks through Earth's normally radio-opaque ionosphere; the lower end of this range is remarkably close to that identified by the authors. These CME events generate the beautiful polar auroras, disrupt our use of the medium-wave radio band, and sometimes perturb the background geomagnetic field at Earth's surface enough to disturb animal navigation.

All known sensory systems in animals are based on cells specialized to convert the stimulus of interest into a coded stream of action potentials that are sent to the brain[8]. If the effects of radio-frequency radiation are real, such cells must exist, but the mystery is in the biophysics. The lack of an enhanced effect at the Larmor frequency, and the low levels of radiation concerned, make it unlikely that a previously proposed mechanism[6] for radiosensing, based on light activation of a cellular protein called cryptochrome, is involved. But some magnetic effects on animals (such as that of a short, sharp magnetic pulse[1]) function through biological magnetite (Fe_3O_4) in tissue — might this also be the radio-wave detector?

If it is, how could such a detection mechanism have arisen? Early animals that had a simple compass patterned along the lines of magnetotactic bacteria would have needed to survive geomagnetic excursions or reversals — periods in which Earth's magnetic field weakened — and natural selection would have favoured individuals with higher cellular volumes of magnetite[3,9]. When the field recovered, animals would have been left with cells that have surprisingly large magnetic moments[9] (Fig. 1). Such cells might then have evolved to serve other functions, such as intensity-based magnetic navigation systems, increasing the amount of magnetite further. With large enough volumes of metallically conductive magnetite in these cells, direct detection of the small electric and magnetic vectors of radio-frequency radiation might have emerged, as Engels and colleagues suggest.

Do the authors' findings have implications for humans? It seems that geomagnetic sensitivity dates back to an early ancestor of animals, and it is clearly present in many extant mammalian species. Human tissues also contain biological magnetite[10]. Many people claim to be bothered by radio transmissions, and some have even moved to live in radio-frequency 'quiet zones' around radio telescopes. Modern-day charlatans will undoubtedly seize on this study as an argument for banning the use of mobile phones, despite the different frequency bands involved. However, if the effect reported by the authors stands the acid test of reproducibility, we might consider gradually abandoning our use of this portion of the electromagnetic spectrum and implementing engineering approaches to minimize incidental low-frequency noise, to help migratory birds find their way. ■

Joseph L. Kirschvink *is in the Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California 91125, USA, and at the Earth-Life Science Institute, Tokyo Institute of Technology, Japan.*
e-mail: kirschvink@gmail.com

1. Wiltschko, W. & Wiltschko, R. *J. Comp. Physiol. A* **191,** 675–693 (2005).
2. Kirschvink, J. L. *Bioelectromagnetics* **13,** 401–411 (1992).
3. Kirschvink, J. L., Winklhofer, M. & Walker, M. M. *J. R. Soc. Interface* **7,** S179–S191 (2010).
4. Kobayashi, A. K., Kirschvink, J. L. & Nesson, M. H. *Nature* **374,** 123 (1995).
5. Engels, S. *et al. Nature* **509,** 353–356 (2014).
6. Ritz, T. *et al. Biophys. J.* **96,** 3451–3457 (2009).
7. LaBelle, J. & Treumann, R. A. *Space Sci. Rev.* **101,** 295–440 (2002).
8. Block, S. M. in *Sensory Transduction* (eds Corey, D. P. & Roper, S. D.) Ch. 1, 1–17 (Rockefeller Univ. Press, 1992).
9. Eder, S. H. K. *et al. Proc. Natl Acad. Sci. USA* **109,** 12022–12027 (2012).
10. Kirschvink, J. L., Kobayashi-Kirschvink, A. & Woodford, B. J. *Proc. Natl Acad. Sci. USA* **89,** 7683–7687 (1992).
11. Schumann, D. *et al. Proc. Natl Acad. Sci. USA* **105,** 17648–17653 (2008).

This article was published online on 7 May 2014.

EVOLUTION

Geology and climate drive diversification

Data from the Galapagos Islands exemplify how geology and climate can interact to cause episodes of isolation and fusion of the biota across a landscape. Different scales of such cycles dictate varying mechanisms of species generation.

ROSEMARY G. GILLESPIE & GEORGE K. RODERICK

Writing in the *Journal of Biogeography*, Ali and Aitchison[1] examine geological and climatic events over the past 700,000 years, namely island ontogeny and shifting sea levels, and their effects on biodiversity in the Galapagos Islands. The authors propose a process that can be considered a general evolutionary mechanism: that the dynamics of isolation caused by geological and climatological processes plays a fundamental part in shaping diversity. Whether these processes promote or constrain species diversification, however, depends on the spatial (global, regional or local) and temporal (multimillion, multimillennial or multidecadal) scales and periodicity of isolation and coalescence.

Geological events have long been known to mould and shape biodiversity. A breakthrough in understanding the underlying mechanisms came with the recognition that ancient splitting of landmasses resulted in shared diversity. The concept of vicariance biogeography — the separation of a group of organisms by a geographical barrier — provided the means for rigorous hypothesis testing in a hitherto largely descriptive field. This established that vicariance resulting directly from geological events can cause diversification, such that geological history will be clearly reflected in the resulting biotic assemblages. The isolation created by ancient geological events is fundamental. Yet, what is given is frequently taken away — separate land masses can become connected and biotic assemblages reunited to various degrees. For example, the Great American

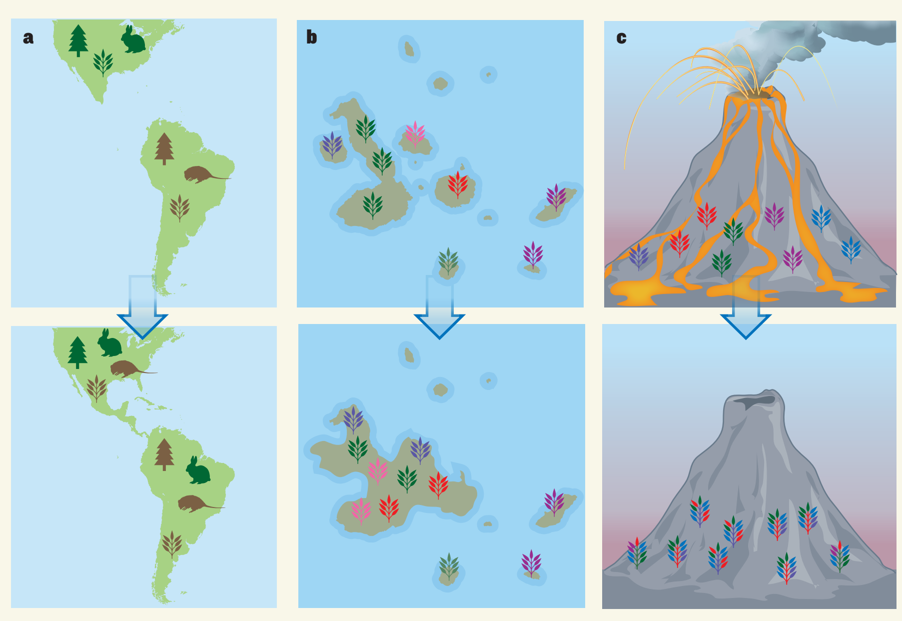


Figure 1 | Isolation across scales of space and time. **a**, On a multimillion-year timescale, distinct assemblages of biota form in isolation on different continents, and joining of the continents may lead to biotic exchange or species displacement. **b**, On a millennial timescale, different species may evolve in isolation on different islands; connections between these islands can then result in a richer biological assemblage as newly formed species come together. **c**, On a decadal timescale, populations may be isolated by recurrent events, such as volcanic eruptions; subsequent reconnection of these populations may result in new genetic combinations, a phenomenon that may also occur with invasive species.

Interchange associated with the formation of the Isthmus of Panama some 3 million years ago allowed the exchange of biotas of North and South America, each of which had evolved in isolation.

There is increasing evidence that geological and climatological events over shorter timescales (10,000 to 1 million years) and over smaller spatial scales (regional or local rather than global) can also influence the diversification process. A prime example of insights to emerge from such situations is provided by Ali and Aitchison in their model, which integrates the dynamics of geological and climatological events that repeatedly connected and disconnected islands of the Galapagos over the past 700,000 years. During this time period, these intermittent connections allowed otherwise landlocked vertebrates to disperse to other islands and to reconnect with populations that had previously been separated. More broadly, Ali and Aitchison use detailed palaeogeographical reconstructions to provide a series of explicit hypotheses about past population structuring and species formation that can be tested with molecular and geospatial data from extant species.

Although repeated separation and reuniting of biotas over a multimillennial time frame can obscure older geological and climatological events, the dynamics of geology and climate can be powerful forces in generating biodiversity. For example, work focused on time frames of 10,000 to 100,000 years ago has shown that climatological events can act as a 'species pump', in which periods of warming or drying serve to alternately isolate and reunite biotas. This phenomenon is well illustrated by Indonesian ants in the Sundaland rainforest, in which diversification has been attributed to repeated episodes of separation and connection of populations during the Plio-Pleistocene (around 5 million to 12,000 years ago), associated with fluctuating sea levels and climate[2]. Similar episodic isolation associated with climatic shifts has been inferred for the diversification of Amazonian vertebrates in the Pleistocene (around 2 million to 12,000 years ago)[3]. Even in the oceans, which have the potential for extensive mixing, opposing processes of isolation and exchange seem to have been responsible for much diversification[4].

On more recent (decadal) timescales and in localized areas, the same mechanism of isolation and mixing resulting from the combined effects of geological and climatological events can influence patterns of diversity prior to the formation of species. For example, population mixing and hybridization — resulting from previously separated populations coming back into contact — have had a key role in generating adaptive variation and functional novelty in populations of cichlid fish in African lakes[5]. Indeed, the repeated isolation and subsequent mixing of populations in new combinations may serve as an 'evolutionary crucible' to facilitate and potentially accelerate diversification[6]. Furthermore, the negative consequences of founder effects (the reduced genetic variation that occurs when a population is established by a small number of individuals) may be offset if different colonization events result in multiple genotypes within the introduced population. This process highlights the potential role of mixing among successively colonizing populations in providing the genetic variation to allow adaptive evolution[7].

Thus, geological and climatological dynamics over time and space shape the patterns of biodiversity that we can observe and measure today. Specifically, the periodicity of isolation and connection dictates evolutionary outcomes, and understanding of this dynamic has become the focus of genomic approaches[8]. When species evolve in isolation over long (multimillion-year) periods, reconnections can result in exchange or displacement of entire assemblages (Fig. 1a). When species are isolated for long enough to allow speciation (multimillennia), subsequent connections may unite newly formed species, thereby generating diversity (Fig. 1b). And when populations are isolated locally and over decadal timescales, thus prompting the development of disparate gene pools, subsequent reconnection and mixture can create new genetic combinations upon which selection can act (Fig. 1c).

Moving to the present and future, it is clear that, as organisms shift their distributions in response to climate change and as globalization increasingly homogenizes previously isolated biotas, understanding the role of historic isolation and recent connection in biodiversity dynamics is crucial. Indeed, the mixing of previously isolated populations is characteristic of many invasive species, suggesting a role for novel genetic combinations in their successful establishment in new environments. An increased knowledge of past events of isolation and fusion of biotas, such as that provided by Ali and Aitchison, will better equip us to predict future dynamics of biodiversity. ■

Rosemary G. Gillespie *and* **George K. Roderick** *are in the Department of Environmental Science and the Berkeley Natural History Museums, University of California Berkeley, Berkeley, California 94720-3114, USA.*
e-mails: gillespie@berkeley.edu; roderick@berkeley.edu

1. Ali, J. R. & Aitchison, J. C. *J. Biogeogr.* http://dx.doi.org/10.1111/jbi.12313 (2014).
2. Quek, S.-P., Davies, S. J., Ashton, P. S., Itino, T. & Pierce, N. E. *Mol. Ecol.* **16,** 2045–2062 (2007).
3. Haffer, J. *Biodivers. Conserv.* **6,** 451–476 (1997).
4. Bowen, B. W., Rocha, L. A., Toonen, R. J., Karl, S. A. & the ToBo Laboratory *Trends Ecol. Evol.* **28,** 359–366 (2013).
5. Loh, Y.-H. E. *et al. Mol. Biol. Evol.* **30,** 906–917 (2013).
6. Carson, H. L., Lockwood, J. P. & Craddock, E. M. *Proc. Natl Acad. Sci. USA* **87,** 7055–7057 (1990).
7. Rius, M. & Darling, J. A. *Trends Ecol. Evol.* **29,** 233–242 (2014).
8. Feder, J. L., Flaxman, S. M., Egan, S. P., Comeault, A. A. & Nosil, P. *Annu. Rev. Ecol. Syst.* **44,** 73–97 (2013).

doi:10.1038/nature13274

Recent advances in homogeneous nickel catalysis

Sarah Z. Tasker[1]*, Eric A. Standley[1]* & Timothy F. Jamison[1]

Tremendous advances have been made in nickel catalysis over the past decade. Several key properties of nickel, such as facile oxidative addition and ready access to multiple oxidation states, have allowed the development of a broad range of innovative reactions. In recent years, these properties have been increasingly understood and used to perform transformations long considered exceptionally challenging. Here we discuss some of the most recent and significant developments in homogeneous nickel catalysis, with an emphasis on both synthetic outcome and mechanism.

To the uninitiated, nickel might seem like just the impoverished younger sibling of palladium in the field of transition metal catalysis. After all, the use of palladium-catalysed cross-coupling has skyrocketed over the past half-century: it was honoured with the 2010 Nobel Prize in Chemistry, and is ubiquitous in applications that range from complex natural product synthesis to drug discovery to manufacturing. Nickel lies just above palladium in the periodic table, and as a group 10 metal, it can readily perform many of the same elementary reactions as palladium or platinum. Because of these commonalities, nickel is often viewed solely as a low-cost replacement catalyst for cross-coupling reactions. However, this common misconception is clearly refuted by the numerous and diverse nickel-catalysed reactions reported in the literature. Indeed, homogeneous nickel catalysis is currently experiencing a period of intensified interest. In this Review, we aim to use recent developments in organonickel chemistry to illustrate how the intrinsic properties of nickel have enabled its use as an effective catalyst for many intriguing, valuable and difficult transformations.

Historically, the use of nickel in organometallic reactions pre-dates many other examples of transition metal catalysis[1,2]. Nickel was isolated in 1751; its name is derived from the German *Kupfernickel*, the name given to a nickel ore originally believed by miners to contain copper, but which did not yield copper on extraction (hence use of *Nickel*, a mischievous demon). In the 1890s, Mond observed one of the unusual reactivity patterns of nickel: elemental nickel and CO reacted at room temperature to form $Ni(CO)_4$, an extremely toxic, low-boiling liquid, which could be used to purify the metal. Shortly thereafter, Sabatier performed the first hydrogenation of ethylene using nickel, for which he was awarded the 1912 Nobel Prize in Chemistry. But undoubtedly, one of the most prominent and prolific early contributors to organonickel chemistry was Wilke[1]. Wilke made seminal contributions to the structure and reactivity of nickel complexes, including the synthesis of $Ni(cod)_2$ (a ubiquitous source of complexed zero-valent nickel, Ni(0); cod, 1,5-cyclooctadiene) and investigation of olefin oligomerization reactions.

Beginning in the 1970s, nickel found extensive use both for cross-coupling and reactions of alkenes and alkynes, such as nucleophilic allylation, oligomerization, cycloisomerization and reductive coupling. Many excellent books and reviews of organonickel chemistry in general[2], as well as of specific transformations (for example, reductive coupling[3] and cross-coupling[4]), already exist. Consequently, we have chosen to focus on key advances in nickel-catalysed reactions since 2005 and to highlight how researchers can take advantage of nickel's characteristic properties and reactivity to perform innovative and useful transformations. Whereas applications of nickel chemistry span materials science, polymer synthesis and biocatalysis, this Review encompasses only homogeneous nickel catalysis relevant to small molecule synthesis. Additionally, owing to the short nature of this Review and the breadth of nickel chemistry, we are unable to include discussions of all the exemplary methods developed in the past decade. However, we hope that the selected reactions and mechanistic studies presented herein spark further investigation into the full range of nickel-catalysed reactions.

Mechanism and elementary steps

Before discussing each class of transformation, a survey of nickel's characteristic modes of reactivity, particularly in regard to some of the elementary steps of transition metal catalysis (Fig. 1) is needed. Nickel is a relatively electropositive late transition metal. Therefore, oxidative addition[5], which results in loss of electron density around nickel, tends to occur quite readily (though, conversely, reductive elimination is correspondingly more difficult)[6]. This facile oxidative addition allows for the use of cross-coupling electrophiles that would be considerably less reactive under palladium catalysis, such as phenol derivatives[7–9], aromatic nitriles[10] or even aryl fluorides[11].

Nickel also has a number of readily available oxidation states commonly invoked in catalysis. The majority of palladium-catalysed reactions are based on a Pd(0)/Pd(II) catalytic cycle, and most often proceed through polar (that is, non-radical) mechanisms. Likewise, Ni(0)/Ni(II) catalytic cycles are widespread, but the easy accessibility of Ni(I) and Ni(III) oxidation states allows different modes of reactivity and radical mechanisms. As a result, many transformations are based on Ni(I)/Ni(III), Ni(0)/Ni(II)/Ni(I), or even cycles in which nickel remains in the Ni(I) state for the entire catalytic cycle[12].

Many nickel complexes have long been known as privileged catalysts for reactions of alkenes and alkynes, such as oligomerization[13] or reductive coupling. Nickel readily donates d-electrons to π-acceptors, so olefin bonding is generally strong[14]. β-Hydride elimination tends to be slower with nickel relative to palladium; specifically, the energy barrier to Ni–C bond rotation prior to β-hydride elimination is often significantly higher for nickel than comparable palladium species[15].

Finally, there are a few more obvious differences between nickel and its group 10 counterparts. Practically speaking, the cost of nickel in its elemental form is roughly 2,000 times lower than palladium and 10,000 times lower than platinum on a mole-for-mole basis, though the price of commonly used nickel sources for catalysis can be less favourable. As a first-row transition metal, nickel has a small atomic radius, and Ni–ligand bond lengths are often relatively short[16]. Researchers have been taking advantage of all of the above features to develop new reactions, which are demonstrated in the specific examples that follow.

[1]Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.
*These authors contributed equally to this work.

a

	Nickel	Palladium
Oxidation states	−1 **0** **+1** **+2** **+3** +4	**0** +1 **+2** +3 **+4**
	Smaller atomic radius	Larger atomic radius
	Less electronegative	More electronegative
	Harder	Softer
	Facile oxidative addition	Facile reductive elimination
	Facile β-migratory insertion	Facile β-hydride elimination
	Radical pathways more accessible	

b

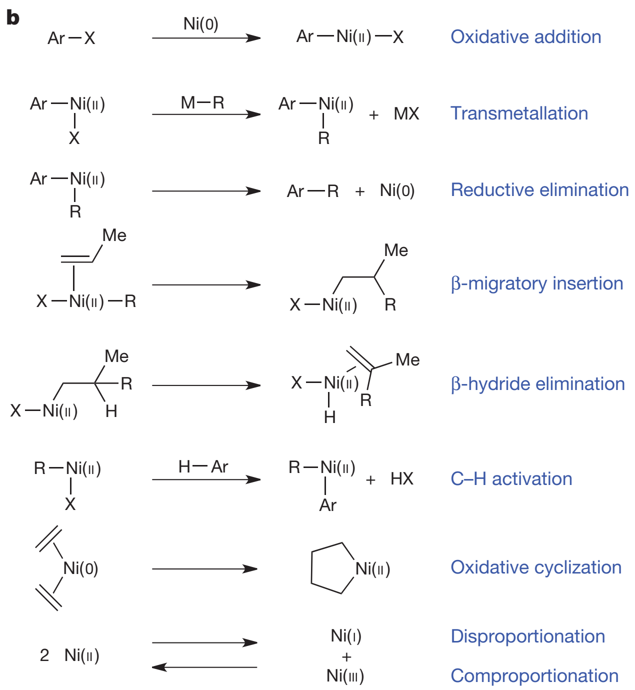


Figure 1 | Nickel fundamentals. **a**, Comparison of basic characteristics of nickel and palladium, including accessible oxidation states (top row, states in bold are more commonly involved in catalysis) and trends in reactivity. **b**, Prototypical examples of elementary organometallic reaction steps, highlighting changes in oxidation state at nickel. Additional ligands bound to nickel not involved in each transformation are omitted for clarity. Ar, aryl; M, metal; Me, methyl.

Cross-coupling

Building on discoveries and developments from the 1970s, nickel has proved to be an extremely effective catalyst for cross-coupling. Cross-coupling reactions are transition-metal-catalysed reactions originally developed as a means to synthesize biaryls from arylmetal species and aryl halides or pseudohalides. Alongside palladium, nickel has been used extensively for Suzuki–Miyaura (organoboron reagents) and Negishi (organozinc reagents) cross-coupling reactions, in particular[17]. In more recent years, the scope of cross-coupling reactions has expanded far beyond simple biaryl synthesis to include many other types of coupling partner. Nickel catalysis has been at the forefront of this expansion, as will be demonstrated in the following sections.

Cross-coupling of aryl halides

The ability of nickel to oxidatively insert into carbon–heteroatom bonds with ease is of particular advantage in the Suzuki–Miyaura reaction of heteroaromatic boronic acids, which readily undergo protodeboronation under the basic (and usually hydrous) reaction conditions, particularly at elevated reaction temperatures. One useful contribution comes from Hartwig and Ge, who have developed a method for the synthesis of heterobiaryls that takes advantage of this rapid oxidative addition by combining a lowered reaction temperature with readily activated precatalyst **3** to afford the traditionally difficult-to-access heterobiaryls[18] (Fig. 2a). Additionally, precatalyst **3** possesses adequate stability to be handled in air, and only 0.5 mol% of the precatalyst is required, making this a highly efficient reaction. This method is applicable to both heteroaryl chlorides and bromides and gives high yields

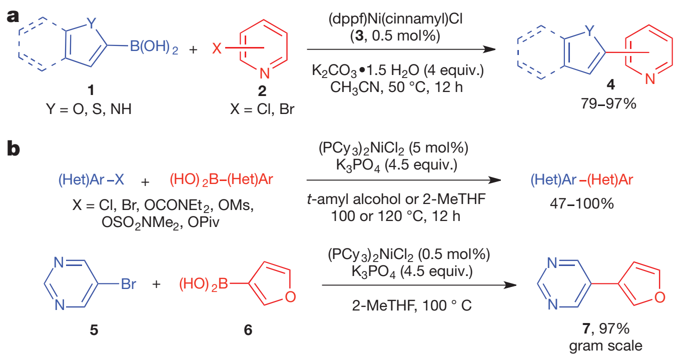


Figure 2 | Recent nickel-catalysed Suzuki–Miyaura arylations. **a**, Cross-coupling of heteroaryl boronic acids (**1**) with heteroaryl halides (**2**) to form heterobiaryls (**4**) is a long-standing challenge. This method, employing an air-stable nickel catalyst precursor **3**, provides the desired heterobiaryls in excellent yields. **b**, Cross-couplings developed for small-scale use are often carried out in solvents poorly suited to industrial or large-scale use. As such, the adaptation of the Suzuki–Miyaura cross-coupling to form (hetero)biaryls such as **7** using 'green' solvents while still obtaining the products in high yield is a valuable development. dppf, 1,1′-bis(diphenylphosphino)ferrocene; cinnamyl, *trans*-$C_6H_5CHCHCH_2$–; (Het)Ar, heteroaryl; THF, tetrahydrofuran; Cy, cyclohexyl; Ms, methanesulphonate; Piv, pivaloyl.

of products across an impressive range of substrates. A related method for the synthesis of heterobiaryls was disclosed by Garg and co-workers in 2013[19] (Fig. 2b). This method focused on further improving the efficiency of Suzuki–Miyaura couplings as well as employing 'green' solvents such as 2-MeTHF and *t*-amyl alcohol.

Cross-coupling of phenol derivatives

Although using aryl halides as cross-coupling partners is the de facto standard, accessing the desired halide coupling partner is not always trivial and can sometimes be extremely challenging. One well-established solution is the use of aryl triflates (Fig. 3a). They typically possess extremely high reactivity in cross-coupling reactions and, because they are obtained by reaction of a phenol with triflic anhydride, are derived from a pool of materials entirely separate from aryl halides. For these two reasons, triflates are valuable coupling partners. However, triflates are prone to hydrolysis, especially under basic conditions, making their use for certain reactions challenging. Tosylates and mesylates (Fig. 3b), close relatives of triflates, have found use in cross-coupling reactions, as their greater stability reduces or eliminates the problem

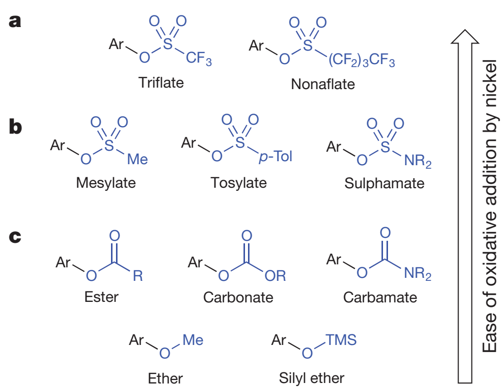


Figure 3 | Halogen alternatives used in cross-coupling reactions. **a**, Aryl triflates have long been used as replacements for halogens in cross-coupling reactions. Aryl nonaflates were developed later to address some of the issues encountered when working with aryl triflates, but their use is less widespread. **b**, The use of aryl mesylates, tosylates and sulphamates presents many advantages over triflates and related fluorinated sulphonates owing to their increased stability. **c**, Like sulphonate derivatives, the use of carboxylic esters, carbonates, carbamates, ethers and silyl ethers can be advantageous in many situations. *p*-Tol, *para*-tolyl (4-methylphenyl); TMS, trimethylsilyl.

of hydrolysis. However, their reactivity towards oxidative addition by metals is also considerably reduced, often leading to the need for harsher reaction conditions. For these reasons, catalyst systems capable of activating the C–O bonds of functional groups other than triflates—such as ethers, esters, carbamates and carbonates (Fig. 3c)—are highly desired. Reactions based on such catalyst systems would represent an ideal combination of ready access to coupling partners that can be mildly and selectively activated by nickel, yet are robust enough to withstand many conditions that would degrade or decompose the analogous triflate.

Nickel has been known to activate C–O bonds since as early as 1977[20], and since this time, a number of advances in this field have been reported. For example, Kocienski and Dixon developed the first effective Ni(0)-catalysed cross-coupling of vinyl carbamates with organomagnesium reagents in 1989[21], and several years later Snieckus and co-workers described the Ni(0)-catalysed cross-coupling of aryl carbamates and organomagnesium reagents in concert with further functionalization by directed *ortho*-metallation[22]. Subsequently, Snieckus also reported the Ni(0)-catalysed cross-coupling of tertiary aryl sulphamates with organomagnesium reagents[23].

Around this time, the field experienced an increased interest in the nickel-mediated activation of these 'inert' C–O bonds. A seminal development in the field came from Dankwardt[24], who described the cross-coupling of aryl ethers (**8**) with arylmagnesium reagents (**9**) in a Kumada–Corriu-type coupling (Fig. 4a). Another critical advance in this area of research came from Chatani and co-workers in 2008, when they reported the first Suzuki–Miyaura cross-coupling using aryl ethers (**8**) and boronic esters (**10**)[25] (Fig. 4b). Previously, all couplings of this type had employed Grignard reagents, which are generally poorly compatible with many common functional groups; the change to boronic esters, however, provided a considerable improvement to the substrate scope and allowed much easier implementation of this chemistry.

In 2008, in nearly simultaneous reports, Shi and co-workers reported the Suzuki–Miyaura cross-coupling of aryl acetates and pivalates with boroxines (**13**)[26], while Garg and co-workers reported the Suzuki–Miyaura cross-coupling of aryl pivalates (**11**) with boronic acids (**12**)[27] (Fig. 3c). Both methods are capable of producing a wide variety of biaryls, demonstrating their viability as valuable alternatives to the use of traditional aryl halides and sulphonate esters. Following these initial reports, it was demonstrated that organozinc reagents[28] and aryl alkoxides[29,30] can be used, further expanding the range of available coupling partners. Additional advances were realized through a collaborative experimental and theoretical investigation carried out by Houk, Sneickus, Garg and co-workers[31].

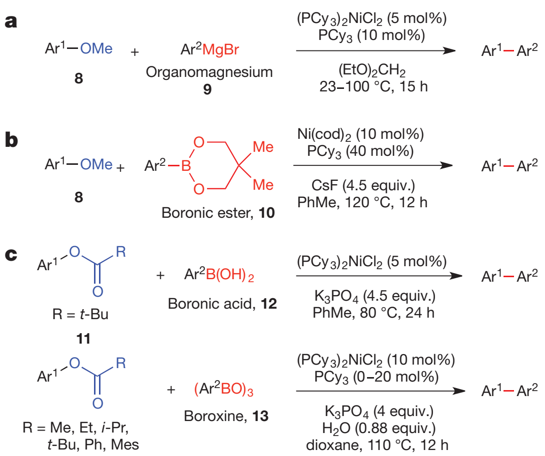


Figure 4 | Milestones in cross-coupling reactions of aryl ethers and esters. **a**, Kumada–Corriu-type, nickel-catalysed biaryl formation from aryl ethers (**8**) and organomagnesium (Grignard) reagents (**9**). **b**, Suzuki–Miyaura-type, nickel-catalysed biaryl synthesis from aryl ethers (**8**) and boronic esters (**10**). **c**, Suzuki–Miyaura-type, nickel-catalysed biaryl synthesis using aryl esters (**11**) and aryl boronic acids (**12**) or aryl boroxines (**13**). Ph, phenyl; *t*-Bu, *tert*-butyl; Et, ethyl; cod, 1,5-cyclooctadiene; *i*-Pr, isopropyl; Mes, 2,4,6-trimethylphenyl.

In the period since these reports, many fruitful developments have been made to allow the transformation of many types of readily accessible phenol derivative into valuable biaryls[32], and other types of reaction beginning from phenol derivatives have since followed, such as the Mizoroki–Heck reaction reported by Watson and co-workers[33]. In addition to C–C bond forming reactions, methods for the amination[34,35] and reduction[12,36] of phenol derivatives have been disclosed.

Benzylic cross-coupling

Another mode of bond activation that has come to prominence, at least in part thanks to nickel catalysis, is the activation of benzylic C–O bonds[37]. Benzylic ethers, esters, carbonates, carbamates and, in some instances, even free alcohols (via the corresponding magnesium alkoxide) can be activated by low-valent nickel. With proper choice of starting material and organometallic reagent, the reaction products can be di- or triarylalkanes, both of which are ubiquitous motifs in drug targets, natural products and materials applications. Perhaps the most significant feature of these transformations is their high stereospecificity, which allows access to these products in highly enantioenriched form from the (readily available) corresponding enantioenriched mono- or diarylmethanol.

In 2011, Jarvo and co-workers disclosed the first stereospecific nickel-catalysed alkyl–alkyl cross-coupling reaction[38] (Fig. 5a). This method, in contrast to previous nickel-catalysed cross-couplings using sp^3 electrophiles, does not racemize the alkyl electrophile and in this instance provides clean inversion of the starting stereochemistry. Therefore, the existing stereochemistry of the starting material (**14**) can be used as the only source of chiral information, rather than relying on catalyst control with or without a directing group that must be later removed from the molecule. Jarvo demonstrated the synthesis of several interesting and useful molecules using this method, including the anti-cancer agent **16** from the corresponding methyl ether. The product was obtained with 96% enantiomeric excess (e.e.) beginning from material of 98% e.e., which demonstrates the high stereospecificity of this method. A subsequent publication demonstrated the use of the 2-methoxyethyl ether moiety as a directing/activating group, which greatly increases the ease with which oxidative addition into the C–O bond takes place[39].

Furthermore, in a subsequent publication, the Jarvo group demonstrated the Suzuki–Miyaura cross-coupling of benzylic esters, carbonates and carbamates with arylboronic esters[40] (Fig. 5b). A striking feature of this method is that it affords retention (**17a**) or inversion (**17b**) of the starting stereochemistry based on the ligand employed—the use of tricyclohexylphosphine (PCy_3) provides retention of stereochemistry, whereas using an *N*-heterocyclic carbene ligand (SIMes; Fig. 5b) provides inversion. In both cases, the enantiospecificity is greater than 97% across a wide range of substrates. Simultaneously with the disclosure from the Jarvo group, Watson and co-workers published a closely related method for the synthesis of diarylalkanes and triarylmethanes beginning from benzylic pivalates (**18**) and boroxines (**13**)[41] (Fig. 5c). In contrast to Jarvo's method, however, it was found that no external phosphine or carbene ligand was necessary to obtain good stereospecificity—$Ni(cod)_2$ alone was demonstrated to be an effective catalyst and provided inversion of configuration at the benzylic position.

In 2012, Shi and co-workers reported the direct cross-coupling of benzylic alcohols (**20**) with Grignard reagents (**21**) to provide diarylmethanes or alkyl arenes with a catalyst system composed of $(PCy_3)_2NiCl_2$ and PCy_3 (ref. 42; Fig. 5d). The magnesium alkoxide, the active coupling partner, is pre-formed by addition of MeMgBr, which is also used to activate the Ni(II) precatalyst by reducing it to Ni(0) via a sequence of two successive transmetallations and reductive elimination of ethane. Subsequent addition of the Grignard reagent initiates the reaction to form the desired coupling product.

Whereas all reactions so far described involve cleavage of C–O bonds, benzylic halides are also useful substrates for this type of activation. One clever example is from Martin and co-workers, who transformed benzylic halides to the corresponding phenylacetic acids using carbon dioxide as the carbon source[43]. Subsequently, this methodology was extended to aryl and benzylic pivalates to synthesize benzoates and phenylacetic acids[44]. Additionally, the activation of benzylic ammonium salts has recently been demonstrated[45].

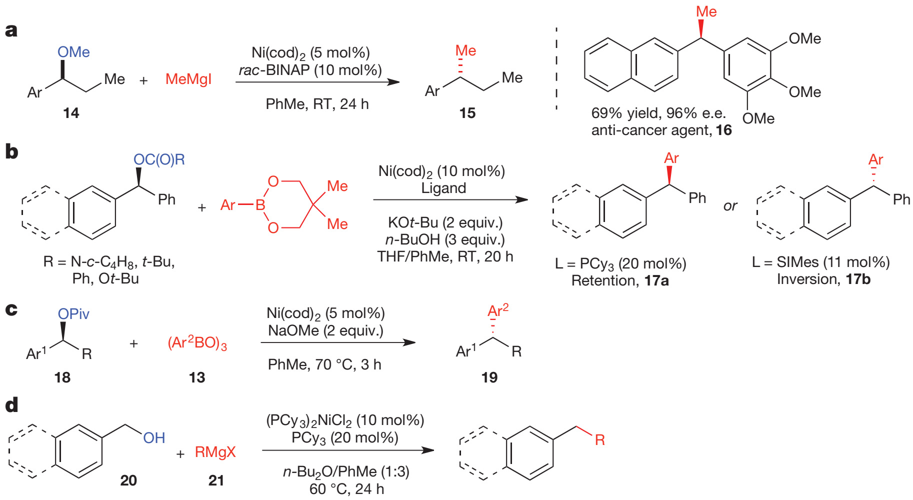


Figure 5 | Reactions of benzylic alcohols and alcohol derivatives. **a**, Stereospecific methylation of benzylic ethers. A nickel catalyst comprising Ni(cod)$_2$ and *rac*-BINAP was found to catalyse the methylation of benzylic methyl ethers (**14**) to form alkyl-substituted arenes (**15**). A modification for the synthesis of diarylethanes was also devised, allowing the synthesis of the anti-cancer agent **16** in 69% yield and 96% e.e. **b**, A Suzuki–Miyaura-type arylation of benzylic esters, carbonates and carbamates. The synthesis of triarylmethanes (**17**) can be achieved by catalytic Ni(cod)$_2$ and PCy$_3$ or SIMes—the stereoselectivity (retention or inversion, respectively **17a** or **17b**) is determined by the identity of the ligand. **c**, A phosphine- and carbene-free nickel catalyst was also developed, yielding inversion of the stereochemistry of benzylic pivalates (**18**) to provide access to diarylalkanes (**19**). **d**, Cross-coupling of free benzylic alcohols. An excess of organomagnesium reagent (**21**) can be added to form a magnesium alkoxide, which is then a competent coupling partner for the Kumada-type coupling with organomagnesium reagents. *rac*, racemic; BINAP, 2,2′-bis(diphenylphosphino)-1,1′-binaphthyl; SIMes, (1,3-bis(2,4,6-trimethylphenyl)-4,5-dihydroimidazol-2-ylidene); Bu, butyl.

Cross-coupling of aziridines

The activation of C–N bonds, specifically those of aziridines, by zero-valent nickel has been known for a number of years[46]. This activation is facile, and, intriguingly, it can be reversed on exposure to molecular oxygen to again afford the original aziridine. However, it was not until a full decade later that this mode of activation was successfully incorporated into a catalytic coupling reaction. Doyle and co-workers succeeded in coupling alkyl and arylzinc halides (**23**) to styrene-derived *N*-tosyl aziridines (**22**) to yield aziridine ring opening with excellent specificity for addition at the more substituted position of the aziridine (**24**)[47] (Fig. 6a). Critical to the success of this approach was the use of dimethyl fumarate as an additive in place of the phosphine, carbene and/or amine ligands often used in nickel catalysts. Dimethyl fumarate is believed to accelerate reductive elimination through π-coordination to nickel. Alkyl-substituted aziridines, unfortunately, were found to be unsuitable substrates for this set of conditions. However, the use of the dual-purpose cinsyl group, which functions both as a removable protecting group and a directing group, enabled the use of aliphatic aziridines (**25**)[48]. In this way, good selectivity (2.5:1 to 4.9:1) for substitution at the less substituted position could be achieved with moderate to high yields (Fig. 6b).

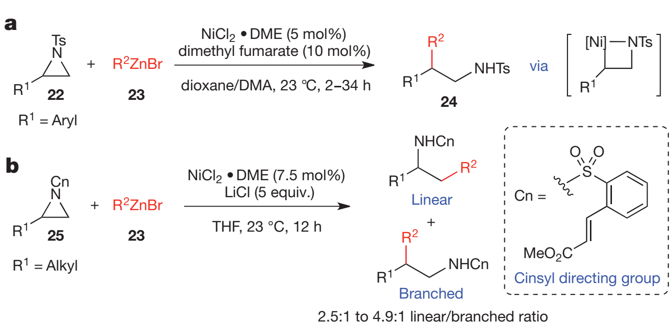


Figure 6 | Nickel-catalysed Negishi-type cross-coupling of aromatic and aliphatic aziridines. **a**, Nickel-catalysed addition of organozinc halides to styrenyl aziridines (**22**) occurs with incorporation of the nucleophile (**23**) at the substituted position of the aziridine to furnish β,β-disubstituted sulphonamides (**24**). **b**, Nickel-catalysed addition of organozinc halides (**23**) to aliphatic aziridines (**25**) directed by the cinsyl group (Cn, shown in dashed box), which imparts a preference for addition at the less substituted position of the aziridine. Ts, (4-methyl)phenylsulphonyl; DME, 1,2-dimethoxyethane; DMA, dimethylacetamide.

Cross-coupling of sp^3 halides

The utility of cross-coupling of aryl or vinyl electrophiles is unquestioned with regard to scope, functional group compatibility, and relevance to the arene-rich pharmacophores ubiquitous in modern drug molecules. However, the number of C(sp^3)–C(sp^3) linkages found in natural products and other complex organic molecules far outstrips the number of arene–carbon linkages. But C(sp^3)–C(sp^3) bond forming reactions, particularly those involving tertiary or quaternary stereocentres, can be challenging, even with modern methods. In the past decade, a large amount of progress has been made towards an ultimate goal of the capability to form aliphatic carbon–carbon bonds stereoselectively at will, just as the formation of biaryl or aryl–heteroatom bonds has already been revolutionized by the field of cross-coupling.

Some key challenges are associated with any cross-coupling in which the electrophile (that is, the component that undergoes oxidative addition to the metal catalyst) is sp^3 hybridized[49,50]. First, the activation energy for oxidative addition can be large, given that C(sp^3)–X bonds are more electron-rich than C(sp^2)–X bonds. For primary electrophiles, oxidative addition can proceed by an S_N2-like inversion pathway, but for secondary or tertiary electrophiles, this pathway is very slow[51]. Then, once the carbon–metal bond has been formed, the challenge is to suppress intramolecular β-hydride elimination, which would produce an alkene product. Transmetallation and/or reductive elimination, therefore, must proceed more rapidly than β-hydride elimination. Cases in which β-hydrogens are not available or not of the correct geometry for elimination, such as in benzyl or adamantyl electrophiles, as discussed previously, are also known but do not represent the same level of difficulty.

Despite these obstacles, reactions of sp^3 electrophiles were investigated in the early days of cross-coupling[52]. Suzuki reported the first Pd- or Ni-catalysed C(sp^3)–C(sp^3) cross-coupling reaction, primary alkyl iodides with alkyl boranes using Pd(PPh$_3$)$_4$, but significant amounts of reduction and elimination products were also formed (reduction:elimination:desired coupling, 27:9:50)[53]. Knochel then showed that nickel could be used to successfully couple primary iodides with organozinc reagents using either an intramolecularly tethered alkene[54] or exogenous electron-poor alkene[55]

to facilitate reductive elimination. Kambe reported an olefin-assisted Kumada coupling of primary alkyl halides and tosylates, proposed to proceed via a bis(η^3-allyl)nickel catalyst formed by the coupling of two equivalents of butadiene[56].

However, the modern era of C(sp^3)–C(sp^3) cross-coupling began with a paper by Fu and co-workers in 2003, in which the coupling of secondary alkyl bromides with β-hydrogens and organozinc reagents was reported[57] (Fig. 7a). The transition from previously used primary electrophiles to secondary electrophiles was ground-breaking, because it opened the way to asymmetric synthesis of tertiary or quaternary all-carbon stereocentres. The chelating tridentate PyBOX ligand (**28**) was found to be essential, probably in order to slow the rate of β-hydride elimination, which would require a vacant coordination site. In addition to the Negishi reaction, the cross-coupling of secondary electrophiles with aryl and vinyl boronic acids[58], aryl silicon reagents[59] and aryl tin reagents[60] has been disclosed. Cross-coupling reactions with a variety of C(sp^3)–metal reagents can also be accomplished

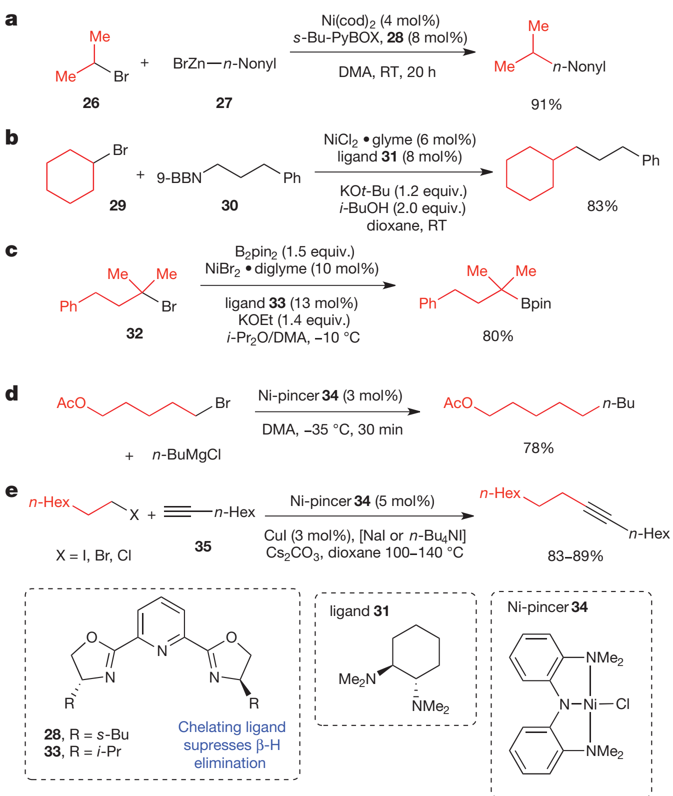


Figure 7 | Key representative examples of cross-coupling reactions involving oxidative addition to sp^3 carbon electrophiles. a, The first example of a C(sp^3)–C(sp^3) cross-coupling reaction using a secondary electrophile. Secondary bromides and iodides, such as **26**, were coupled in a Negishi reaction with primary alkyl zinc reagents, such as **27**. It was proposed the chelating PyBOX ligand (**28**) blocks the open coordination site needed for undesired β-hydride elimination. **b**, A mild (room temperature) Suzuki reaction of secondary bromides (**29**) and primary alkylboranes (**30**). Previously used bipyridyl or PyBOX ligands were unable to promote the transformation, so diamino ligand **31** was used. **c**, The first cross-coupling reaction to use an unactivated tertiary electrophile (**32**). In contrast to previous results, tertiary electrophiles reacted with faster rates than secondary or primary electrophiles. **d**, The Kumada coupling of primary alkyl bromides and iodides, and some secondary alkyl iodides, with Grignard reagents was accomplished with the Ni-pincer complex **34** with an amidobis(amine) ligand. Low temperatures allow a wide range of functional groups (such as ketones and esters) to be tolerated. **e**, The first example of a Ni-catalysed Sonogashira reaction of alkyl electrophiles. β-Hydrogen-containing alkyl iodides, bromides and chlorides could be used as the electrophile with a variety of terminal alkynes (**35**). 9-BBN, 9-borabicyclo[3.3.1]nonane; pin, pinacolato; glyme, bis(2-methoxyethyl) ether; Pr, propyl; Hex, hexyl.

using this chemistry; for example, the mild Suzuki coupling carried out at room temperature shown in Fig. 7b[61]. Finally, Fu and co-workers provided the first examples of the cross-coupling of a tertiary alkyl electrophile (for example, **32**) with B_2pin_2 (pin, pinacolato; Fig. 7c)[62] and later in a Suzuki reaction[63]. In stark contrast, palladium has only been reported to cross-couple secondary sp^3 electrophiles in a handful of specialized cases[64].

Preliminary mechanistic investigations of these reactions are consistent with an inner-sphere electron-transfer pathway for oxidative addition. For these coupling reactions, Fu proposes[62,63] a Ni(I)/Ni(III) cycle with a radical oxidative addition pathway (Fig. 8a), although the Vicic group has demonstrated[65] that with redox-active ligands such as terpyridine, the transmetallated species before oxidative addition is perhaps better thought of as Ni(II) cationic species **39** bound to reduced radical anion ligands (Fig. 8a, in square brackets). However, the first isolable and well-characterized Ni(III) complex arising from a Ni(I)/Ni(III) oxidative addition was also recently reported[66]. In any case, these mechanistic manifolds clearly demonstrate the ability of Ni to access various oxidation states in order to facilitate reactions otherwise inaccessible with group 10 metals.

Another prominent contributor to the field of C(sp^3)–C(sp^3) cross-coupling has been Hu, who developed Ni-pincer complex **34**[67]. This complex has proven catalytically active for C(sp^3)–C(sp^3) Kumada cross-couplings[68] (Fig. 7d) as well as the Sonogashira coupling of alkynes with primary alkyl iodides, bromides and chlorides[69] (Fig. 7e). Extensive mechanistic investigation has been conducted with the former reaction (Fig. 8b), revealing a number of interesting features[70]. The active complex for the turnover-limiting step of transmetallation appears to be the Ni-pincer complex coordinated to the Grignard reagent (**34′**). A bimetallic oxidative addition of the primary alkyl halide occurs via generation of an alkyl radical (**43**), which reacts with a different Ni(II)-alkyl complex (**41**), and then undergoes reductive elimination to form the product. The remaining unstable Ni(I) complex (**45**) transfers an electron to a Ni(III) complex (**42**) to form two Ni(II) complexes (**41**, **46**), which can rejoin the catalytic cycle.

These examples of cross-coupling are impressive, but the promise of secondary C(sp^3)–C(sp^3) cross-coupling lies in asymmetric reactions providing tertiary and quaternary stereocentres with high enantioselectivity. The Fu group has published numerous examples of this type of reactivity. These approaches rely on a tethered directing group that can coordinate to the nickel catalyst on oxidative addition to form a rigid complex such that, on reductive elimination, enantioenriched products are generated (Fig. 9). Because the electrophile probably undergoes a radical oxidative addition, both enantiomers of the starting alkyl halide are converted through a common planar radical intermediate. Therefore, a racemic alkyl halide can be used to produce a highly enantioenriched product.

The first example of an asymmetric secondary C(sp^3)–C(sp^3) cross-coupling was published in 2005: a Negishi reaction of α-bromo amides (**47**) with organozinc reagents[71] (Fig. 9a). The use of an (R)-PyBOX ligand (**33**) provided good yields and excellent enantioselectivities (>90% e.e. in nearly all cases). These conditions not only conferred good enantioselectivity, but also provided selectivity for oxidative addition to the α-bromide in the presence of other primary or secondary alkyl bromides.

Although amides are useful building blocks for further synthetic elaboration, the need for a directing group does limit the generality of this approach. The Fu group has shown that a wide variety of directing groups can be used (Fig. 9b). Not only do various coordinating functional groups such as amides, esters, sulphones and alkenes work well in these reactions, but the tether length can vary as well. In fact, the Fu group has shown that an amide in conjunction with the chiral nickel catalyst can confer high levels of enantioselectivity even when the halide is in the δ-position, a full five atoms away from the amide nitrogen[72]. Doubtless, as it develops, the asymmetric cross-coupling reaction will become more and more general and be used for challenging bond formation in the synthesis of natural products and other more complex molecules[73,74].

Reductive cross-coupling

Traditional cross-coupling reactions join electrophilic (for example, aryl bromide) and nucleophilic (for example, aryl zinc) components. However,

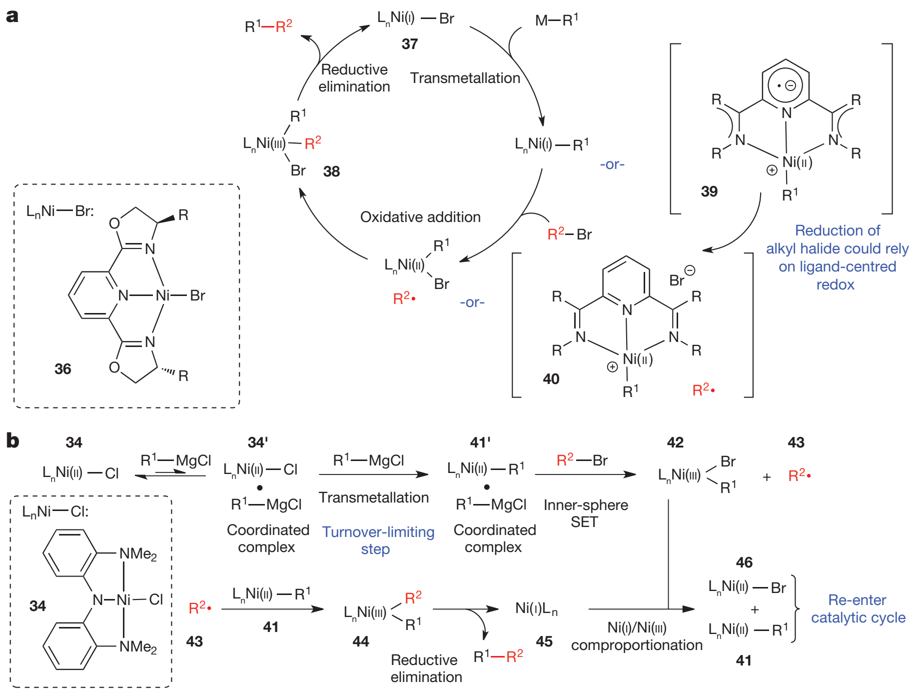


Figure 8 | Proposed mechanisms for C(sp^3)–C(sp^3) cross-coupling. **a**, Mechanism proposed for Fu-type cross-coupling reactions in which ligands are typically PyBOX or similar as shown (**36**). A Ni(I) complex (**37**) undergoes transmetallation, then a radical oxidative addition of the electrophile, to eventually form a Ni(III) complex (**38**), which can reductively eliminate the coupled product. Nickel complexes with redox active ligands have been shown to be perhaps better thought of as the species in brackets (**39**, **40**), in which the oxidative addition to the electrophile proceeds through ligand-centred rather than metal-centred redox. Ni(II) compounds are used as precatalysts, allowing for reactions to be set up outside the glove box. Reduction to Ni(I) presumably occurs before the beginning of the catalytic cycle via reduction of Ni(II) to Ni(0) (transmetallation/reductive elimination), then comproportionation of Ni(0)/Ni(II). **b**, Mechanism proposed for the C(sp^3)–C(sp^3) Kumada cross-coupling with Ni-pincer **34**. Extensive mechanistic studies have shown that a more complex bimetallic oxidative addition is operative, in which a Ni(II) complex bound to an equivalent of Grignard reagent (**34′**) is the active complex for the turnover-limiting transmetallation. SET, single electron transfer; R^2•, alkyl radical; • indicates a coordination complex.

nucleophilic reagents can be unstable, costly or cause undesired side reactions. In contrast, reductive cross-coupling reactions join two electrophilic components without the formation of a stoichiometric organometallic species. The challenge inherent in reductive cross-coupling processes is the differentiation between the two electrophiles in order to suppress homodimerization of either component.

Recently, Weix and co-workers demonstrated an impressive solution to these difficulties by performing the first catalytic reductive cross-coupling to show high selectivities for cross-coupled products over dimerization without resorting to the use of a large excess of one of the components[75] (Fig. 10a). By coupling an aryl iodide (**49**) and an alkyl iodide (**50**) electrophile with $NiI_2 \cdot xH_2O$ and both a bipyridyl and a phosphine ligand, as well as a stoichiometric manganese reducing agent, high reactivity and selectivity for cross-coupled products were observed. Later, a range of aryl bromides and chlorides and alkyl bromides were shown to be reactive with zinc as the reducing agent[76]. Though Zn(0) is present, no organozinc species are formed; rather, zinc (or manganese) acts as a reductant directly to the nickel centre, accounting for the excellent functional group compatibility.

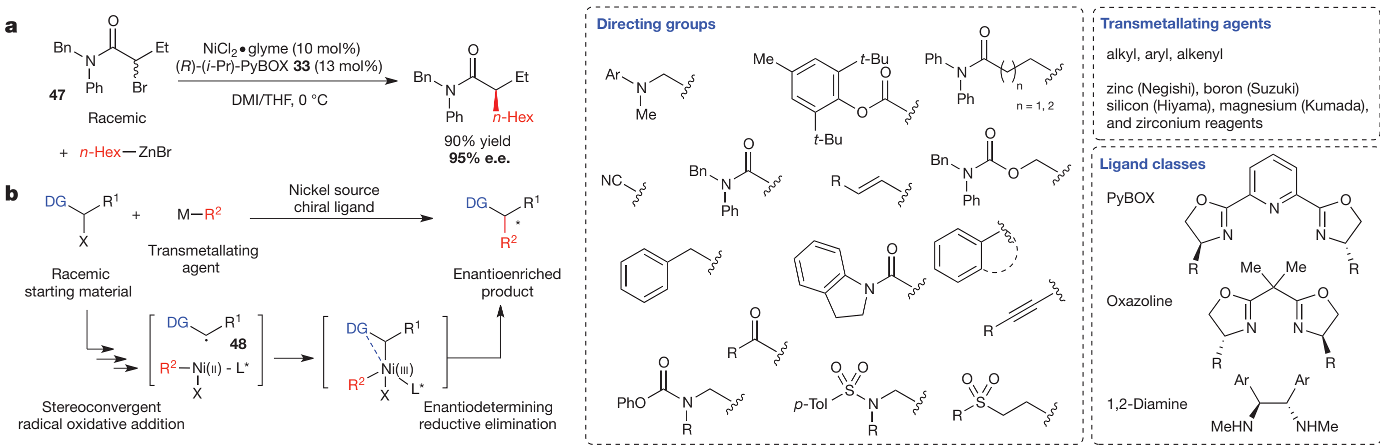


Figure 9 | Asymmetric C(sp^3)–C(sp^3) cross-coupling reactions. **a**, The first example of an asymmetric C(sp^3)–C(sp^3) cross-coupling between a racemic α-bromoamide (**47**) and aliphatic organozinc reagent. **b**, Proposed mechanism for the coupling reaction. A racemic aliphatic halide adjacent to a directing group (in large dashed box) undergoes oxidative addition. Since it is proposed to proceed via an unligated aliphatic radical (**48**) as shown in Fig. 8a, the overall oxidative addition is stereoconvergent, and thus the chiral ligand on nickel dictates the stereochemistry of the product. Many classes of directing groups and transmetallating reagents (dashed box at top right) have been successfully reacted, generally relying on one of the three chiral ligand classes shown in the dashed box at bottom right. Bn, benzyl; DMI, 1,3-dimethyl-2-imidazolidinone.

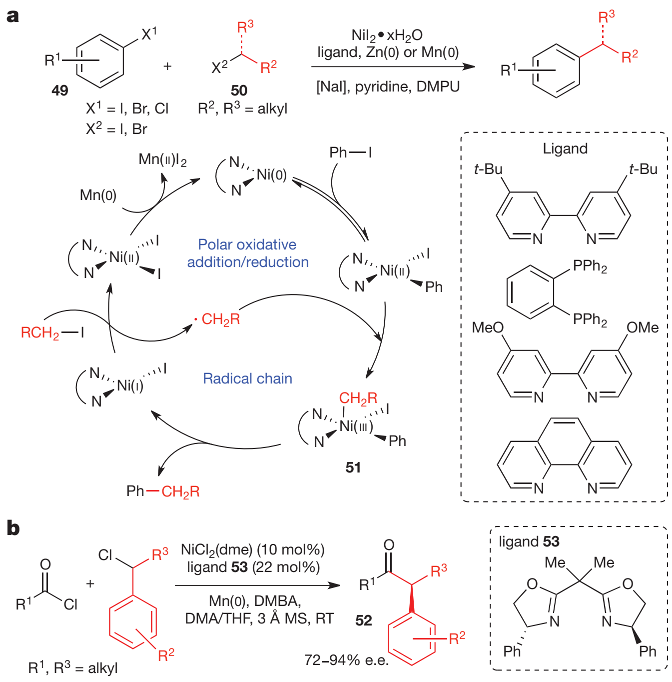


Figure 10 | Reductive cross-coupling reactions. **a**, The reductive cross-coupling reaction of an aryl halide (**49**) with an alkyl halide (**50**) without the intermediacy of an organozinc or organomanganese species. Extensive mechanistic studies have suggested that this method combines both polar (aryl halide) and radical chain (alkyl halide) formal oxidative addition mechanisms. Because the oxidation state of nickel is matched to each electrophile, homodimerization is suppressed. For details on possible methods of radical chain initiation, see ref. 77. **b**, First asymmetric acyl reductive cross-coupling. High enantioselectivities are obtained with bisoxazoline ligands such as **53**. DMPU, 1,3-dimethyl-3,4,5,6-tetrahydro-2(1*H*)-pyrimidinone; DMBA, 2,6-dimethylbenzoic acid; MS, molecular sieves.

After extensive mechanistic investigation, Weix proposed[77] the mechanism in Fig. 10a. Nickel undergoes a polar oxidative addition to the aryl halide followed by a radical chain generation of an alkyl radical. This radical then undergoes coordination to form Ni(III) species **51** competent for reductive elimination. In this way, the full potential of readily available nickel oxidation states (see above) is realized, and the use of alkyl electrophiles is allowed. Additionally, the properties of the halides are matched to the capabilities of the catalyst to avoid side reactions: aryl halides more readily undergo polar oxidative addition (Ni(0)/Ni(II)), and alkyl radicals are more stable than aryl radicals (Ni(I)/Ni(III)). Close comparison of this mechanism with the one shown in Fig. 8b also demonstrates that a number of complex elementary steps are available with radical pathways, and the factors that lead to certain patterns of reactivity have not been fully understood.

The Reisman group has also developed an enantioselective reductive coupling that affords α,α-disubstituted ketones (**52**) from the coupling of acyl chlorides and secondary benzylic chlorides[78] (Fig. 10b). Although extensive mechanistic investigations have not been carried out, Reisman also proposes a mechanism in which Mn(0) acts as a direct reductant to Ni. In any case, further developments of asymmetric reductive cross-coupling would be a valuable alternative to the use of traditional transmetallating agents for cross-coupling.

C–H activation

C–H activation, the direct functionalization of a hydrocarbon, presents several potential advantages over traditional cross-coupling in that prefunctionalization of the substrate, for example with a halogen, is not necessary. At the same time, however, controlling site-selectivity with C–H activation can be very challenging, as molecules often contain many C–H bonds with similar chemical properties. Though examples of nickel-mediated C–H activation date to at least 1963[79], the development of catalytic C–H activation methods using nickel are more recent.

One such example, from the research group of Itami, describes the decarbonylative coupling of aryl and heteroaryl esters (**55**) and (benz)oxazoles (**54**)[80,81] (Fig. 11a). This reaction provides an unusual means of arylating

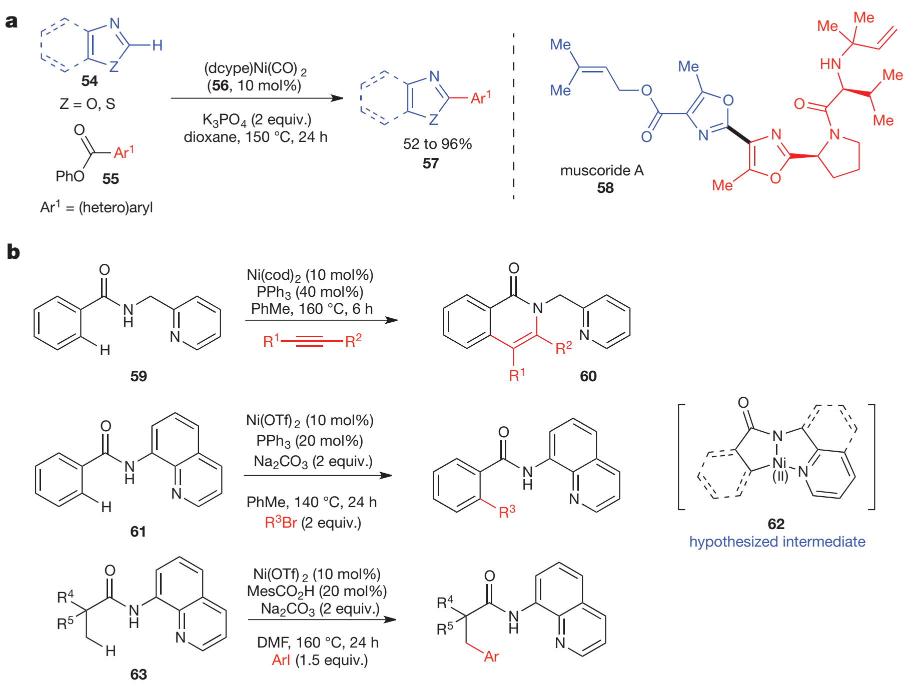


Figure 11 | Selected examples of nickel-catalysed C–H activation reactions. **a**, Benzoxazoles and benzothiazoles (**54**) are useful substrates for this nickel-catalysed C–H activation reaction, which uses aryl esters (**55**) as the electrophilic coupling partners to produce (hetero)biaryls (**57**). This methodology was applied to the formal synthesis of muscoride A (**58**) to great effect. **b**, Nickel-catalysed, chelation-assisted C–H activation reactions have recently been developed. These reactions rely on a directing group to facilitate addition of nickel into the C–H bond in the *ortho* position of a benzamide (**59**, **61**) or into the C–H bond of an adjacent aliphatic substituent (**63**). dcype, 1,2-bis(dicyclohexylphosphino)ethane; OTf, triflate (trifluoromethanesulphonate); DMF, dimethylformamide.

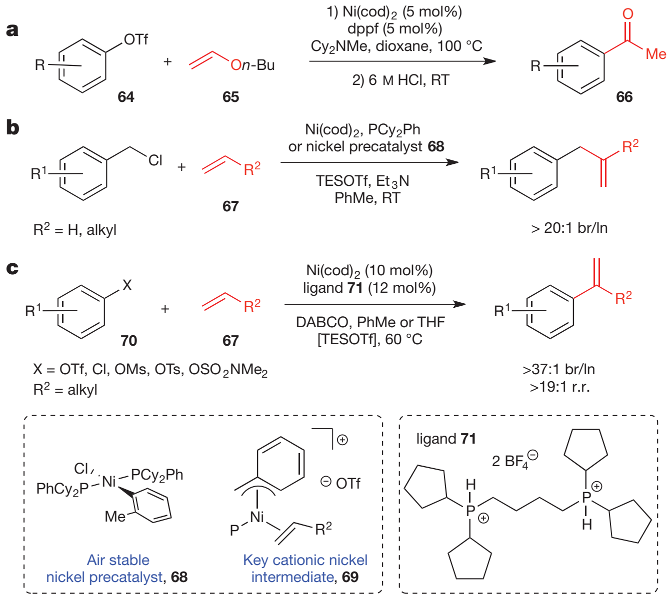


Figure 12 | Nickel-catalysed Heck reactions. a, Coupling of aryl triflates (**64**) with electron-rich enol ethers (**65**) to obtain high selectivity for branched products, which on acidic hydrolysis form ketones. Computational work supports a cationic Heck pathway with catalyst regeneration as turnover-limiting. **b**, The first Heck reaction highly selective for branched (br) products with electronically unbiased (aliphatic) and non-chelating alkenes (**67**). Again, it was proposed to proceed through a cationic Ni species (**69**; left dashed box) to give high regioselectivity, and an air-stable precatalyst (**68**; left dashed box) was developed to eliminate the need for an air-free technique. **c**, Branch-selective Heck reaction for aryl electrophiles (**70**) with aliphatic olefins (**67**). Bidentate ligand **71** (right dashed box) was key to both reactivity of aryl electrophiles and suppression of undesired isomerization. Aryl chlorides and other phenol-derived electrophiles can be used with the addition of TESOTf, which is proposed to perform a counterion exchange in order to enter the cationic Heck pathway. TESOTf, triethylsilyl trifluoromethanesulphonate; br/ln, branched-to-linear product ratio; DABCO, 1,4-diazabicyclo[2.2.2]octane; r.r., regioisomeric ratio—the ratio of desired product to all other isomers.

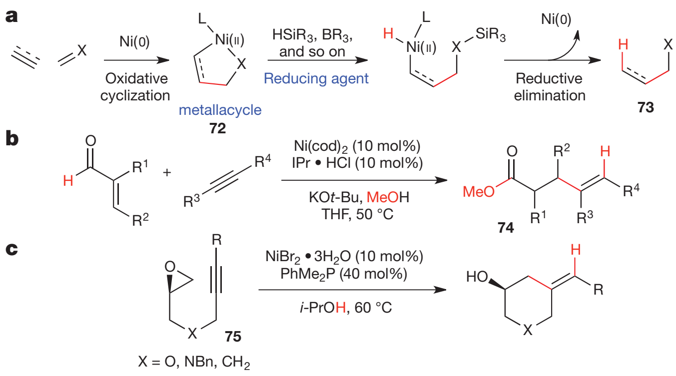


Figure 13 | Prototypical reductive coupling reactions and use of new reducing agents. a, Standard reductive coupling reaction. Oxidative cyclization of two π-components forms a nickellacycle (**72**), which on formation of a Ni–H bond with a reducing agent, undergoes reductive elimination to form a new C–C σ-bond and a new C–H bond overall (**73**). **b**, Use of methanol as a mild reducing agent via the intermediacy of a hemiacetal. **c**, Use of isopropanol as a mild, external reducing agent, allowing for the use of air-stable Ni(II) salts as precatalysts. IPr, 1,3-bis(2,6-diisopropylphenyl)-1,3-dihydro-2*H*-imidazol-2-ylidene.

azoles, and adds to the existing decarbonylative/decarboxylative approaches for C–H arylation. The yields of the isolated products are good to excellent across a wide range of substituted esters and azoles, and, furthermore, the air-stable precatalyst $Ni(dcype)(CO)_2$ (**56**) can be used in place of $Ni(cod)_2$ and dcype, making this reaction operationally simple (dcype, 1,2-bis(dicyclohexylphosphino)ethane). Itami used this method to execute an expedient formal synthesis of muscoride A (**58**) in excellent yield. Furthermore, a subsequent report from Itami and co-workers describes the isolation of the key arylnickel(II) pivalate intermediate formed in couplings of this type[82].

In 2011, Chatani and co-workers reported the first example of a nickel-catalysed, chelation-assisted C–H activation reaction[83]. The reaction, an oxidative cycloaddition of alkynes to aromatic amides to form 1-isoquinolones **60** (Fig. 11b), relies on a 2-pyridylmethyl group on the amide nitrogen to function as a directing group. Subsequently, in 2013, Chatani and Tobisu published the *ortho* C–H activation of a similar aryl amide system[84] (**61**) and the arylation of aliphatic C–H substituents[85] (**62**), further expanding the repertoire of nickel-catalysed C–H activation reactions. All three reactions are hypothesized to occur by pre-coordination of nickel to the 1,2-diamine moiety of the directing group, followed by metallation to form **62**, after which the reaction paths diverge.

Heck reaction

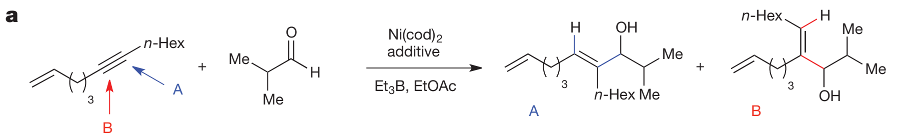
Many of the advantages of nickel catalysis can also be applied to the Heck reaction[86]. The Heck reaction is similar to cross-coupling reactions, but rather than undergo transmetallation, the oxidative addition complex coordinates an alkene. Subsequent migratory insertion and β-hydride elimination then furnishes a more substituted alkene. Computational work comparing Ni and Pd for use in the Heck reaction has been carried out, with Guo and co-workers proposing lower energy barriers for oxidative addition and migratory insertion for Ni in contrast to faster β-hydride elimination and catalyst regeneration for Pd (ref. 15). However, relatively little has been done to truly capitalize on these features until quite recently.

Questions of regioselectivity are critical in the Heck reaction, because in theory migratory insertion can occur at either end of the alkene coupling partner. Traditionally, electron-poor alkenes such as styrenes and acrylates have been used, which confer high selectivity for addition to the terminal position of the olefin. Conditions under which a cationic Pd or Ni species is formed by dissociation of the halide or triflate component, rather than the phosphine, after oxidative addition can provide high selectivity for electron-rich alkenes[87]. An example of the latter using Ni catalysis is given by Skrydstrup and co-workers, who demonstrated good selectivity for coupling of aryl triflates (**64**) and enol ethers (**65**), which following subsequent hydrolysis formed methyl ketones (**66**)[88] (Fig. 12a).

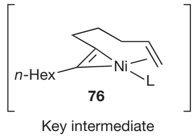
However, electronically unbiased olefins had always given mixtures of branched and linear product isomers arising from addition of the electrophile to the internal and terminal positions of the alkene, respectively. In 2011, Jamison and co-workers reported the first highly selective Heck reaction for these olefins including ethylene and terminal aliphatic alkenes (**67**)[89] (Fig. 12b). This reaction is proposed to proceed via a cationic Heck pathway; it is suggested that the shorter Ni–ligand bond lengths make steric differentiation between the H and alkyl substituents of the alkene feasible. Later, an air-stable nickel precatalyst (**68**) was developed for this reaction, which obviates the need for $Ni(cod)_2$ (and therefore the use of a glove box) and increases reaction rates[90]. Finally, a branch-selective Heck reaction of the more typically used aryl electrophiles (**70**) with electronically unbiased olefins has recently been reported[91] (Fig. 12c). A wide range of electrophiles including aryl triflates, chlorides and other less reactive sulphonates (mesylates, tosylates, sulphamates) can be used, underscoring the ability of nickel to undergo oxidative addition to a broad range of cheap, stable, traditionally unreactive electrophiles.

Reductive coupling

In addition to cross-coupling, one of the prototypical nickel-catalysed reactions is reductive coupling. Reductive coupling involves the joining of two π-components with a reducing agent to form a new σ-bond between the coupling partners and a new C–H σ-bond arising from the reducing agent (Fig. 13a). The reaction is generally accepted to proceed via a concerted oxidative cyclization, followed by either coordination of the reducing agent to Ni or σ-bond metathesis to provide a nickel hydride, and terminated by



Additive	A : B	Comments
None	> 95 : 5	Alkene stays bound, directs aldehyde to distal position (A)
PBu_3	50 : 50	Two phosphines bind, displacing alkene, aldehyde can displace either
$PCyp_3$	5 : > 95	Aldehyde displaces alkene ligand at proximal position (B) rather than phosphine



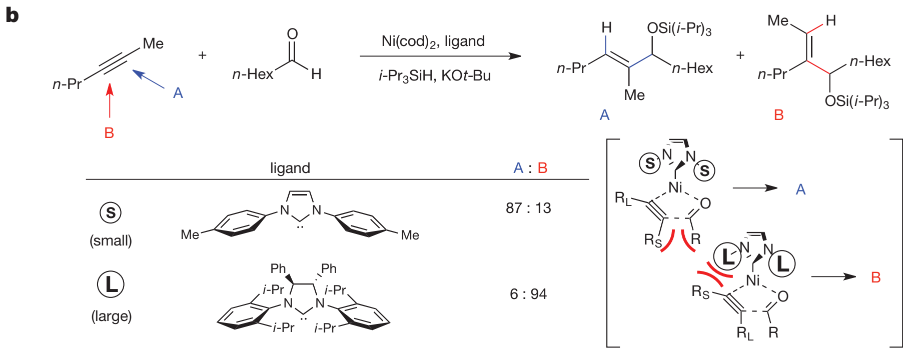


Figure 14 | Two strategies for regiocontrol in reductive coupling reactions. **a**, A tethered alkene can be used to control regioselectivity in coupling an alkyne with an aldehyde. The alkyne oxidatively adds to Ni to produce nickellacyclopropene **76**. Then, by selective displacement by added phosphine ligands, or with no additive, the binding and orientation of the aldehyde is controlled to produce either linear (A) or branched (B) products. **b**, The steric profile of carbene ligands (small substituents, S, compared to large substituents, L) can be used to control the regioselectivity of an alkyne–aldehyde reductive coupling. Computational work suggests that unfavourable steric interactions between the groups on the alkyne with either the group on the aldehyde or the groups on the ligand dictate the orientation of the forming five-membered nickellacycle intermediate (shown in brackets). EtOAc, ethyl acetate; Cyp, cyclopentyl.

reductive elimination. Several excellent reviews[3] of the state of the field before 2005 exist including couplings of both alkenes[92] and alkynes[93].

In more recent years, there have been two key developments in this field. The first development involves the search for milder reducing agents. Traditionally, reactive hydride donors such as alkyl silanes or trialkyl boranes have been used. However, in 2008, Montgomery and co-workers reported that methanol could be used to facilitate β-hydride elimination and provide the nickel hydride directly from an aldehyde in internal redox to form **74**[94] (Fig. 13b). A few years later, Jamison and Beaver reported that by using alcohols directly as the source of the requisite hydride, air-stable Ni(II) salts (reduced under the reaction conditions) could be used to carry out reductive couplings between epoxides and tethered alkynes (**75**)[95] (Fig. 13c).

The second development has addressed a major challenge inherent in reductive coupling reactions: the question of regioselectivity (Fig. 14). For strongly electronically biased coupling components, such as aryl–alkyl alkynes, regioselectivity is controlled by these elements. However, dialkyl alkynes pose a much tougher problem. The first high levels of regiocontrol were achieved by the use of alkynes with a pendant alkene, which could act as a ligand to Ni after coordination to the alkyne[96] (Fig. 14a). Then, by adjusting the ligand sphere around Ni by the addition of phosphine ligands, either position of the alkyne (**A** or **B** in Fig. 14a) could undergo reaction with the aldehyde[97]. Montgomery and co-workers subsequently discovered that by alteration of the steric bulk of carbene ligand substituents, the selectivity in the reductive coupling of internal dialkyl alkynes and aldehydes could be reversed[98] (Fig. 14b). Subsequent computational studies suggested that the orientation of the groups bound to nitrogen was important, particularly in filling the quadrant proximal to the alkyne below the plane of the Ni–ligand bond[99]. These impressive results demonstrate the fine-tuned control over these systems that can now be achieved.

Looking forward

Although the field of nickel catalysis has rapidly expanded over the last decade, there are many challenges that remain to be overcome. Through extensive mechanistic studies, including those described above, researchers now understand a great deal more about the elementary steps and oxidation states of nickel in a number of varying reaction manifolds. In many instances, the picture presented is more complex than originally envisioned, often due to the easy access of nickel to multiple oxidation states, and thus catalytic pathways. With this increased understanding, we expect future efforts to be directed towards the design of new catalysts and the development of new transformations that accomplish even more complex bond-forming reactions. The activation of simple electrophiles for cross-coupling, such as C–H bonds or phenol derivatives, the formation of bonds difficult to access with current methodology, such as carbon–fluorine bonds, and the use of coupling partners currently considered challenging, such as carbon dioxide (CO_2), are all areas that will benefit from further development and mechanistic understanding.

In addition, we expect to see further developments in the area of $C(sp^3)$–$C(sp^3)$ bond formation, particularly in expansion of substrate scope and application to the synthesis of complex molecules. Although there are a few examples of the cross-coupling of sp^3 electrophiles used in total synthesis, the adjustment of synthetic strategy has not perhaps yet occurred to enable their use as a foundational tool in organic chemistry. Another conspicuous absence is a general method for the Heck reaction of sp^3 electrophiles[100]. This reaction is particularly challenging because β-H elimination of the electrophile before coupling must be suppressed, but in order to form the desired alkene product, the measures taken to do so must not prohibit a β-H elimination after C–C bond formation.

Finally, we expect great strides in the development of low-cost, air-stable, and easier-to-handle sources of nickel for catalysis. The use of the most common Ni(0) source, $Ni(cod)_2$, requires the use of a glove box, and although in some reactions inexpensive nickel halide sources can be used with subsequent reduction accomplished *in situ*, the development of different modes of activation for nickel precatalysts could lead to wider adoption of nickel catalysis, in both academic and industrial laboratories. All in all, we expect that nickel catalysis will prove a fertile field of study well into the future as chemists continue to address even more challenging problems of reactivity and rapid assembly of complex molecules. We also hope that nickel will continue to gain recognition, not as an inexpensive substitute for palladium, but rather as possessing a number of inherent properties that provide a complement to catalysis by other metals.

Received 17 October 2013; accepted 18 March 2014.

1. Wilke, G. Contributions to organo-nickel chemistry. *Angew. Chem. Int. Edn Engl.* **27,** 185–206 (1988).
2. Tamaru, Y. (ed.) *Modern Organonickel Chemistry* (Wiley-VCH, 2005).
An excellent discussion of nickel catalysis in general.
3. Montgomery, J. Nickel-catalyzed reductive cyclizations and couplings. *Angew. Chem. Int. Edn* **43,** 3890–3908 (2004).
A comprehensive review of reductive coupling before 2004.
4. Diederich, F. & Stang, P. J. (eds) *Metal-Catalyzed Cross-Coupling Reactions* (Wiley-VCH, 1998).
5. Tsou, T. T. & Kochi, J. K. Mechanism of oxidative addition. Reaction of nickel(0) complexes with aromatic halides. *J. Am. Chem. Soc.* **101,** 6319–6332 (1979).
An excellent investigation into the mechanism of oxidative addition of Ni(0) into aryl halides and the factors determining rate, selectivity and solvent effects of such reactions.
6. Lanni, E. L. & McNeil, A. J. Mechanistic studies on Ni(dppe)Cl_2-catalyzed polymerizations: evidence for rate-determining reductive elimination. *J. Am. Chem. Soc.* **131,** 16573–16579 (2009).
7. Li, B.-J., Yu, D.-G., Sun, C.-L. & Shi, Z.-J. Activation of ''inert'' alkenyl/aryl C–O bond and its application in cross-coupling reactions. *Chemistry* **17,** 1728–1759 (2011).
A detailed account of cross-coupling reactions of phenol derivatives of all types, including those catalysed by metals other than nickel.
8. Rosen, B. M. *et al.* Nickel-catalyzed cross-couplings involving carbon–oxygen bonds. *Chem. Rev.* **111,** 1346–1416 (2011).
9. Mesganaw, T. & Garg, N. K. Ni- and Fe-catalyzed cross-coupling reactions of phenol derivatives. *Org. Process Res. Dev.* **17,** 29–39 (2013).
10. Garcia, J. J., Brunkan, N. M. & Jones, W. D. Cleavage of carbon–carbon bonds in aromatic nitriles using nickel(0). *J. Am. Chem. Soc.* **124,** 9547–9555 (2002).
11. Tobisu, M., Xu, T., Shimasaki, T. & Chatani, N. Nickel-catalyzed Suzuki–Miyaura reaction of aryl fluorides. *J. Am. Chem. Soc.* **133,** 19505–19511 (2011).
12. Cornella, J., Gómez-Bengoa, E. & Martin, R. Combined experimental and theoretical study on the reductive cleavage of inert C–O bonds with silanes: ruling out a classical Ni(0)/Ni(II) catalytic couple and evidence for Ni(I) intermediates. *J. Am. Chem. Soc.* **135,** 1997–2009 (2013).
A rigorous experimental and computational study of the hypothesized nickel(I)-based catalytic cycle.
13. O'Connor, C. T. & Kojima, M. Alkene oligomerization. *Catal. Today* **6,** 329–349 (1990).
14. Massera, C. & Frenking, G. Energy partitioning analysis of the bonding in $L_2TM–C_2H_2$ and $L_2TM–C_2H_4$ (TM = Ni, Pd, Pt; L_2 = $(PH_3)_2$, $(PMe_3)_2$, $H_2PCH_2PH_2$, $H_2P(CH_2)_2PH_2$). *Organometallics* **22,** 2758–2765 (2003).
15. Lin, B.-L. *et al.* Comparing nickel- and palladium-catalyzed Heck reactions. *Organometallics* **23,** 2114–2123 (2004).
16. Cordero, B. *et al.* Covalent radii revisited. *Dalton Trans.* 2832–2838 (2008).
17. Yamaguchi, J., Muto, K. & Itami, K. Recent progress in nickel-catalyzed biaryl coupling. *Eur. J. Org. Chem.* 19–30 (2013).
A concise overview of the state of biaryl synthesis by a variety of nickel-catalysed methods.
18. Ge, S. & Hartwig, J. F. Highly reactive, single-component nickel catalyst precursor for Suzuki–Miyuara cross-coupling of heteroaryl boronic acids with heteroaryl halides. *Angew. Chem. Int. Edn* **51,** 12837–12841 (2012).
19. Ramgren, S. D., Hie, L., Ye, Y. & Garg, N. K. Nickel-catalyzed Suzuki–Miyaura couplings in green solvents. *Org. Lett.* **15,** 3950–3953 (2013).
20. Chuit, C., Felkin, H., Frajerman, C., Roussi, G. & Swierczewski, G. Action des organomagnesiens sur les alcools allyliques en presence de complexes du nickel: I. Synthese d'olefines. *J. Organomet. Chem.* **127,** 371–384 (1977).
21. Kocienski, P. & Dixon, N. J. Stereoselective synthesis of homoallylic alcohols by migratory insertion reactions of higher-order cyanocuprates and nickel-catalysed coupling reactions involving enol carbamates. *Synlett* **1989,** 52–54 (1989).
22. Sengupta, S., Leite, M., Raslan, D. S., Quesnelle, C. & Snieckus, V. Nickel(0)-catalyzed cross coupling of aryl O-carbamates and aryl triflates with Grignard reagents. Directed ortho metalation-aligned synthetic methods for polysubstituted aromatics via a 1,2-dipole equivalent. *J. Org. Chem.* **57,** 4066–4068 (1992).
23. Milburn, R. R. & Snieckus, V. The tertiary sulfonamide as a latent directed-metalation group: Ni^0-catalyzed reductive cleavage and cross-coupling reactions of aryl sulfonamides with Grignard reagents. *Angew. Chem. Int. Edn* **43,** 888–891 (2004).
24. Dankwardt, J. W. Nickel-catalyzed cross-coupling of aryl Grignard reagents with aromatic alkyl ethers: an efficient synthesis of unsymmetrical biaryls. *Angew. Chem. Int. Edn* **43,** 2428–2432 (2004).
25. Tobisu, M., Shimasaki, T. & Chatani, N. Nickel-catalyzed cross-coupling of aryl methyl ethers with aryl boronic esters. *Angew. Chem. Int. Edn* **47,** 4866–4869 (2008).
26. Guan, B.-T., Wang, Y., Li, B.-J., Yu, D.-G. & Shi, Z.-J. Biaryl construction via Ni-catalyzed C–O activation of phenolic carboxylates. *J. Am. Chem. Soc.* **130,** 14468–14470 (2008).
27. Quasdorf, K. W., Tian, X. & Garg, N. K. Cross-coupling reactions of aryl pivalates with boronic acids. *J. Am. Chem. Soc.* **130,** 14422–14423 (2008).
28. Li, B.-J. *et al.* Cross-coupling of aryl/alkenyl pivalates with organozinc reagents through nickel-catalyzed C–O bond activation under mild reaction conditions. *Angew. Chem. Int. Edn* **47,** 10124–10127 (2008).
29. Yu, D.-G. *et al.* Direct application of phenolic salts to nickel-catalyzed cross-coupling reactions with aryl Grignard reagents. *Angew. Chem. Int. Edn* **49,** 4566–4570 (2010).
30. Yu, D.-G. & Shi, Z.-J. Mutual activation: Suzuki–Miyaura coupling through direct cleavage of the sp^2 C–O bond of naphtholate. *Angew. Chem. Int. Edn* **50,** 7097–7100 (2011).
31. Quasdorf, K. W. *et al.* Suzuki–Miyaura cross-coupling of aryl carbamates and sulfamates: experimental and computational studies. *J. Am. Chem. Soc.* **133,** 6352–6363 (2011).
32. Antoft-Finch, A., Blackburn, T. & Snieckus, V. *N,N*-diethyl *O*-carbamate: directed metalation group and orthogonal Suzuki–Miyaura cross-coupling partner. *J. Am. Chem. Soc.* **131,** 17750–17752 (2009).
33. Ehle, A. R., Zhou, Q. & Watson, M. P. Nickel(0)-catalyzed Heck cross-coupling via activation of aryl C–OPiv bonds. *Org. Lett.* **14,** 1202–1205 (2012).
34. Tobisu, M., Yasutome, A., Yamakawa, K., Shimasaki, T. & Chatani, N. Ni(0)/NHC-catalyzed amination of *N*-heteroaryl methyl ethers through the cleavage of carbon–oxygen bonds. *Tetrahedron* **68,** 5157–5161 (2012).
35. Hie, L., Ramgren, S. D., Mesganaw, T. & Garg, N. K. Nickel-catalyzed amination of aryl sulfamates and carbamates using an air-stable precatalyst. *Org. Lett.* **14,** 4182–4185 (2012).
36. Álvarez-Bercedo, P. & Martin, R. Ni-catalyzed reduction of inert C–O bonds: a new strategy for using aryl ethers as easily removable directing groups. *J. Am. Chem. Soc.* **132,** 17352–17353 (2010).
37. Guan, B.-T. *et al.* Direct benzylic alkylation via Ni-catalyzed selective benzylic sp^3 C–O activation. *J. Am. Chem. Soc.* **130,** 3268–3269 (2008).
38. Taylor, B. L. H., Swift, E. C., Waetzig, J. D. & Jarvo, E. R. Stereospecific nickel-catalyzed cross-coupling reactions of alkyl ethers: enantioselective synthesis of diarylethanes. *J. Am. Chem. Soc.* **133,** 389–391 (2011).
39. Greene, M. A., Yonova, I. M., Williams, F. J. & Jarvo, E. R. Traceless directing group for stereospecific nickel-catalyzed alkyl–alkyl cross-coupling reactions. *Org. Lett.* **14,** 4293–4296 (2012).
40. Harris, M. R., Hanna, L. E., Green, M. A., Moore, C. E. & Jarvo, E. R. Retention or inversion in stereospecific nickel-catalyzed cross-coupling of benzylic carbamates with arylboronic esters: control of absolute stereochemistry with an achiral catalyst. *J. Am. Chem. Soc.* **135,** 3303–3306 (2013).
41. Zhou, Q., Srinivas, H. D., Dasgupta, S. & Watson, M. P. Nickel-catalyzed cross-couplings of benzylic pivalates with arylboroxines: stereospecific formation of diarylalkanes and triarylmethanes. *J. Am. Chem. Soc.* **135,** 3307–3310 (2013).
42. Yu, D.-G. *et al.* Direct arylation/alkylation/magnesiation of benzyl alcohols in the presence of Grignard reagents via Ni-, Fe-, or Co-catalyzed sp^3 C–O bond activation. *J. Am. Chem. Soc.* **134,** 14638–14641 (2012).
43. León, T., Correa, A. & Martin, R. Ni-catalyzed direct carboxylation of benzyl halides with CO_2. *J. Am. Chem. Soc.* **135,** 1221–1224 (2013).
44. Correa, A., León, T. & Martin, R. Ni-catalyzed carboxylation of C(sp^2)– and C(sp^3)–O bonds with CO_2. *J. Am. Chem. Soc.* **136,** 1062–1069 (2014).
45. Maity, P., Shacklady-McAtee, D. M., Yap, G. P. A., Sirianni, E. R. & Watson, M. P. Nickel-catalyzed cross couplings of benzylic ammonium salts and boronic acids: stereospecific formation of diarylethanes via C–N bond activation. *J. Am. Chem. Soc.* **135,** 280–285 (2013).
46. Lin, B. L., Clough, C. R. & Hillhouse, G. L. Interactions of aziridines with nickel complexes: oxidative-addition and reductive-elimination reactions that break and make C–N bonds. *J. Am. Chem. Soc.* **124,** 2890–2891 (2002).
47. Huang, C.-Y. & Doyle, A. G. Nickel-catalyzed Negishi alkylations of styrenyl aziridines. *J. Am. Chem. Soc.* **134,** 9541–9544 (2012).
48. Nielsen, D. K., Huang, C.-Y. & Doyle, A. G. Directed nickel-catalyzed Negishi cross coupling of alkyl aziridines. *J. Am. Chem. Soc.* **135,** 13605–13609 (2013).
49. Frisch, A. C. & Beller, M. Catalysts for cross-coupling reactions with non-activated alkyl halides. *Angew. Chem. Int. Edn* **44,** 674–688 (2005).
50. Hu, X. Nickel-catalyzed cross coupling of non-activated alkyl halides: a mechanistic perspective. *Chem. Sci.* **2,** 1867–1886 (2011).
51. Hills, I. D., Netherton, M. R. & Fu, G. C. Toward an improved understanding of the unusual reactivity of Pd^0/trialkylphosphane catalysts in cross-couplings of alkyl electrophiles: quantifying the factors that determine the rate of oxidative addition. *Angew. Chem. Int. Edn* **42,** 5749–5752 (2003).
52. Netherton, M. R. & Fu, G. C. Nickel-catalyzed cross-couplings of unactivated alkyl halides and pseudohalides with organometallic compounds. *Adv. Synth. Catal.* **346,** 1525–1532 (2004).
53. Ishiyama, T., Abe, S., Miyaura, N. & Suzuki, A. Palladium-catalyzed alkyl-alkyl cross-coupling reaction of 9-alkyl-9-BBN derivatives with iodoalkanes possessing β-hydrogens. *Chem. Lett.* **21,** 691–694 (1992).
54. Devasagayaraj, A., Stüdemann, T. & Knochel, P. A new nickel-catalyzed cross-coupling reaction between sp^3 carbon centers. *Angew. Chem. Int. Edn Engl.* **34,** 2723–2725 (1996).
55. Giovannini, R. & Knochel, P. Ni(II)-catalyzed cross-coupling between polyfunctionalized arylzinc derivatives and primary alkyl iodides. *J. Am. Chem. Soc.* **120,** 11186–11187 (1998).
56. Terao, J., Watanabe, H., Ikumi, A., Kuniyasu, H. & Kambe, N. Nickel-catalyzed cross-coupling reaction of Grignard reagents with alkyl halides and tosylates: remarkable effect of 1,3-butadienes. *J. Am. Chem. Soc.* **124,** 4222–4223 (2002).
57. Zhou, J. & Fu, G. C. Cross-couplings of unactivated secondary alkyl halides: room-temperature nickel-catalyzed Negishi reactions of alkyl bromides and iodides. *J. Am. Chem. Soc.* **125,** 14726–14727 (2003).
Reports the first secondary C(sp^3)–C(sp^3) cross-coupling.
58. Zhou, J. & Fu, G. C. Suzuki cross-couplings of unactivated secondary alkyl bromides and iodides. *J. Am. Chem. Soc.* **126,** 1340–1341 (2004).

59. Powell, D. A. & Fu, G. C. Nickel-catalyzed cross-couplings of organosilicon reagents with unactivated secondary alkyl bromides. *J. Am. Chem. Soc.* **126,** 7788–7789 (2004).
60. Powell, D. A., Maki, T. & Fu, G. C. Stille cross-couplings of unactivated secondary alkyl halides using monoorganotin reagents. *J. Am. Chem. Soc.* **127,** 510–511 (2005).
61. Saito, B. & Fu, G. C. Alkyl–alkyl Suzuki cross-couplings of unactivated secondary alkyl halides at room temperature. *J. Am. Chem. Soc.* **129,** 9602–9603 (2007).
62. Dudnik, A. S. & Fu, G. C. Nickel-catalyzed coupling reactions of alkyl electrophiles, including unactivated tertiary halides, to generate carbon–boron bonds. *J. Am. Chem. Soc.* **134,** 10693–10697 (2012).
63. Zultanski, S. L. & Fu, G. C. Nickel-catalyzed carbon–carbon bond-forming reactions of unactivated tertiary alkyl halides: Suzuki arylations. *J. Am. Chem. Soc.* **135,** 624–627 (2013).
64. Rudolph, A. & Lautens, M. Secondary alkyl halides in transition-metal-catalyzed cross-coupling reactions. *Angew. Chem. Int. Edn* **48,** 2656–2670 (2009).
65. Jones, G. D. *et al.* Ligand redox effects in the synthesis, electronic structure, and reactivity of an alkyl–alkyl cross-coupling catalyst. *J. Am. Chem. Soc.* **128,** 13175–13183 (2006).
66. Lipschutz, M. I., Yang, X., Chatterjee, R. & Tilley, T. D. A structurally rigid bis(amido) ligand framework in low-coordinate Ni(I), Ni(II), and Ni(III) analogues provides access to a Ni(III) methyl complex via oxidative addition. *J. Am. Chem. Soc.* **135,** 15298–15301 (2013).
67. Csok, Z., Vechorkin, O., Harkins, S. B., Scopelliti, R. & Hu, X. Nickel complexes of a pincer NN_2 ligand: multiple carbon–chloride activation of CH_2Cl_2 and $CHCl_3$ leads to selective carbon–carbon bond formation. *J. Am. Chem. Soc.* **130,** 8156–8157 (2008).
68. Vechorkin, O. & Hu, X. Nickel-catalyzed cross-coupling of non-activated and functionalized alkyl halides with alkyl Grignard reagents. *Angew. Chem. Int. Edn* **48,** 2937–2940 (2009).
69. Vechorkin, O., Barmaz, D., Proust, V. & Hu, X. Ni-catalyzed Sonogashira coupling of nonactivated alkyl halides: orthogonal functionalization of alkyl iodides, bromides, and chlorides. *J. Am. Chem. Soc.* **131,** 12078–12079 (2009).
70. Breitenfeld, J., Ruiz, J., Wodrich, M. D. & Hu, X. Bimetallic oxidative addition involving radical intermediates in nickel-catalyzed alkyl–alkyl Kumada coupling reactions. *J. Am. Chem. Soc.* **135,** 12004–12012 (2013).
71. Fischer, C. & Fu, G. C. Asymmetric nickel-catalyzed Negishi cross-couplings of secondary α-bromo amides with organozinc reagents. *J. Am. Chem. Soc.* **127,** 4594–4595 (2005).
72. Zultanski, S. L. & Fu, G. C. Catalytic asymmetric γ-alkylation of carbonyl compounds via stereoconvergent Suzuki cross-couplings. *J. Am. Chem. Soc.* **133,** 15362–15364 (2011).
73. Gong, H. & Gagné, M. R. Diastereoselective Ni-catalyzed Negishi cross-coupling approach to saturated, fully oxygenated *C*-alkyl and *C*-aryl glycosides. *J. Am. Chem. Soc.* **130,** 12177–12183 (2008).
74. Son, S. & Fu, G. C. Nickel-catalyzed asymmetric Negishi cross-couplings of secondary allylic chlorides with alkylzincs. *J. Am. Chem. Soc.* **130,** 2756–2757 (2008).
75. Everson, D. A., Shrestha, R. & Weix, D. J. Nickel-catalyzed reductive cross-coupling of aryl halides with alkyl halides. *J. Am. Chem. Soc.* **132,** 920–921 (2010).
76. Everson, D. A., Jones, B. A. & Weix, D. J. Replacing conventional carbon nucleophiles with electrophiles: nickel-catalyzed reductive alkylation of aryl bromides and chlorides. *J. Am. Chem. Soc.* **134,** 6146–6159 (2012).
77. Biswas, S. & Weix, D. J. Mechanism and selectivity in nickel-catalyzed cross-electrophile coupling of aryl halides with alkyl halides. *J. Am. Chem. Soc.* **135,** 16192–16197 (2013).
78. Cherney, A. H., Kadunce, N. T. & Reisman, S. E. Catalytic asymmetric reductive acyl cross-coupling: synthesis of enantioenriched acyclic α,α-disubstituted ketones. *J. Am. Chem. Soc.* **135,** 7442–7445 (2013).
79. Kleiman, J. P. & Dubeck, M. The preparation of cyclopentadienyl [*o*-(phenylazo)phenyl]nickel. *J. Am. Chem. Soc.* **85,** 1544–1545 (1963).
80. Muto, K., Yamaguchi, J. & Itami, K. Nickel-catalyzed C–H/C–O coupling of azoles with phenol derivatives. *J. Am. Chem. Soc.* **134,** 169–172 (2012).
81. Amaike, K., Muto, K., Yamaguchi, J. & Itami, K. Decarbonylative C–H coupling of azoles and aryl esters: unprecedented nickel catalysis and application to the synthesis of muscoride A. *J. Am. Chem. Soc.* **134,** 13573–13576 (2012).
82. Muto, K., Yamaguchi, J., Lei, A. & Itami, K. Isolation, structure, and reactivity of an arylnickel(II) pivalate complex in catalytic C–H/C–O biaryl coupling. *J. Am. Chem. Soc.* **135,** 16384–16387 (2013).
83. Shiota, H., Ano, Y., Aihara, Y., Fukumoto, Y. & Chatani, N. Nickel-catalyzed chelation-assisted transformations involving ortho C–H bond activation: regioselective oxidative cycloaddition of aromatic amides to alkynes. *J. Am. Chem. Soc.* **133,** 14952–14955 (2011).
84. Aihara, Y. & Chatani, N. Nickel-catalyzed direct alkylation of C–H bonds in benzamides and acrylamides with functionalized alkyl halides via bidentate-chelation assistance. *J. Am. Chem. Soc.* **135,** 5308–5311 (2013).
85. Aihara, Y. & Chatani, N. Nickel-catalyzed direct arylation of $C(sp^3)$–H bonds in aliphatic amides via bidentate-chelation assistance. *J. Am. Chem. Soc.* **136,** 898–901 (2014).
86. Oestreich, M. (ed.) *The Mizoroki–Heck Reaction* (Wiley, 2009).
87. Cabri, W. & Candiani, I. Recent developments and new perspectives in the Heck reaction. *Acc. Chem. Res.* **28,** 2–7 (1995).
88. Gøgsig, T. M. *et al.* Mild and efficient nickel-catalyzed Heck reactions with electron-rich olefins. *J. Am. Chem. Soc.* **134,** 443–452 (2012).
89. Matsubara, R., Gutierrez, A. C. & Jamison, T. F. Nickel-catalyzed Heck-type reactions of benzyl chlorides and simple olefins. *J. Am. Chem. Soc.* **133,** 19020–19023 (2011).
90. Standley, E. A. & Jamison, T. F. Simplifying nickel(0) catalysis: an air-stable nickel precatalyst for the internally selective benzylation of terminal alkenes. *J. Am. Chem. Soc.* **135,** 1585–1592 (2013).
91. Tasker, S. Z., Gutierrez, A. C. & Jamison, T. F. Nickel-catalyzed Mizoroki–Heck reaction of aryl sulfonates and chlorides with electronically unbiased terminal olefins: high selectivity for branched products. *Angew. Chem. Int. Edn* **53,** 1858–1861 (2014).
92. Ng, S.-Z., Ho, C.-Y., Schleicher, K. D. & Jamison, T. F. Nickel-catalyzed coupling reactions of alkenes. *Pure Appl. Chem.* **80,** 929–939 (2008).
93. Moslin, R. M., Miller-Moslin, K. & Jamison, T. F. Regioselectivity and enantioselectivity in nickel-catalysed reductive coupling reactions of alkynes. *Chem. Commun.* 4441–4449 (2007).
94. Herath, A., Li, W. & Montgomery, J. Fully intermolecular nickel-catalyzed three-component couplings via internal redox. *J. Am. Chem. Soc.* **130,** 469–471 (2008).
95. Beaver, M. G. & Jamison, T. F. Ni(II) salts and 2-propanol effect catalytic reductive coupling of epoxides and alkynes. *Org. Lett.* **13,** 4140–4143 (2011).
96. Miller, K. M. & Jamison, T. F. Ligand-switchable directing effects of tethered alkenes in nickel-catalyzed additions to alkynes. *J. Am. Chem. Soc.* **126,** 15342–15343 (2004).
97. Moslin, R. M. & Jamison, T. F. Mechanistic implications of nickel-catalyzed reductive coupling of aldehydes and chiral 1,6-enynes. *Org. Lett.* **8,** 455–458 (2006).
98. Malik, H. A., Sormunen, G. J. & Montgomery, J. A general strategy for regiocontrol in nickel-catalyzed reductive couplings of aldehydes and alkynes. *J. Am. Chem. Soc.* **132,** 6304–6305 (2010).
99. Liu, P., Montgomery, J. & Houk, K. N. Ligand steric contours to understand the effects of *N*-heterocyclic carbene ligands on the reversal of regioselectivity in Ni-catalyzed reductive couplings of alkynes and aldehydes. *J. Am. Chem. Soc.* **133,** 6956–6959 (2011).
100. Firmansjah, L. & Fu, G. C. Intramolecular Heck reactions of unactivated alkyl halides. *J. Am. Chem. Soc.* **129,** 11340–11341 (2007).

Acknowledgements This work was supported by the NIGMS (GM62755) and the NSF (Graduate Research Fellowship to S.Z.T. and E.A.S.).

Author Contributions S.Z.T. and E.A.S. worked together to outline and write the content, and T.F.J. helped edit the manuscript, references and figures.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence should be addressed to T.F.J. (tfj@mit.edu).

CORRECTIONS & AMENDMENTS

ERRATUM

doi:10.1038/nature13463

Erratum: Recent advances in homogeneous nickel catalysis

Sarah Z. Tasker, Eric A. Standley & Timothy F. Jamison

Nature **509,** 299–309 (2014); doi:10.1038/nature13274

In the first paragraph of this Review, the words 'such as facile oxidative addition and ready access to multiple oxidation states' were inadvertently repeated in the print version. The paper is correct online.

REVIEW

doi:10.1038/nature13085

Nucleotide signalling during inflammation

Marco Idzko[1], Davide Ferrari[2] & Holger K. Eltzschig[3]

Inflammatory conditions are associated with the extracellular release of nucleotides, particularly ATP. In the extracellular compartment, ATP predominantly functions as a signalling molecule through the activation of purinergic P2 receptors. Metabotropic P2Y receptors are G-protein-coupled, whereas ionotropic P2X receptors are ATP-gated ion channels. Here we discuss how signalling events through P2 receptors alter the outcomes of inflammatory or infectious diseases. Recent studies implicate a role for P2X/P2Y signalling in mounting appropriate inflammatory responses critical for host defence against invading pathogens or tumours. Conversely, P2X/P2Y signalling can promote chronic inflammation during ischaemia and reperfusion injury, inflammatory bowel disease or acute and chronic diseases of the lungs. Although nucleotide signalling has been used clinically in patients before, research indicates an expanding field of opportunities for specifically targeting individual P2 receptors for the treatment of inflammatory or infectious diseases.

Nucleotides—particularly ATP—are well known for their function as a universal energy currency[1]. Interestingly, ATP has a completely different role in the extracellular compartment, where it functions as a signalling molecule through the activation of nucleotide receptors[1]. These receptors are referred to as purinergic P2 receptors. In contrast to P1 receptors, which are activated by the ATP metabolite adenosine, P2 receptors are activated by ATP and/or other nucleotides (for example, UTP). On the basis of their signalling properties, P2 receptors can be further subdivided into metabotropic P2Y receptors (P2YRs) that are G-protein-coupled, and ionotropic P2X receptors (P2XRs) that are nucleotide-gated ion channels[2]. Although P2 receptors were originally described on the basis of their functional role in the central nervous system[3,4], more recent studies demonstrate their widespread expression throughout different tissues (Supplementary Table 1) and implicate them in innate or adaptive immune responses[2,5].

Cellular ATP release during inflammatory conditions

During certain conditions—for example inflammatory, ischaemic and hypoxic—several cell types release ATP from intracellular storage pools into the extracellular compartment[2,5,6]. Although ATP release can occur in an uncontrolled fashion (for example, during necrosis), many studies have examined molecular pathways that control extracellular ATP release[5]. For example, inflammatory cells can release ATP via pannexins or connexin hemichannels[2,7]. Pannexins—transmembrane protein channels that connect the intracellular with the extracellular space—have been implicated in the release of ATP from apoptotic cells[8], and other studies have implicated connexins in extracellular nucleotide release[6,7]. Connexins were originally described as gap junction proteins consisting of two hemichannels. However, isolated hemichannels (connexons) can function as conduits between the cytoplasm and the extracellular space, thereby controlling ATP release, for example from inflammatory cells[7] or vascular endothelia[9]. Other studies found that the release of uridine nucleotides such as UTP, UDP and UDP-glucose are increased during cystic fibrosis[10]. Together, these studies indicate that inflammatory disease conditions are associated with the extracellular release of nucleotides.

Molecular structure and signalling cascade of P2YR

P2YR belongs to the G-protein-coupled receptor (GPCR) family and contains an extracellular amino terminus, an intracellular carboxy terminus and seven transmembrane-spanning motifs (Fig. 1). At present, eight distinct mammalian P2YRs have been cloned and characterized ($P2Y_{1/2/4/6/11/12/13/14}R$). The missing numbers represent either non-mammalian receptors ($P2Y_3R$ is the chicken orthologue of human $P2Y_6R$) or other GPCRs that share some sequence homology with P2YRs but cannot be activated by nucleotides (for example, lysophosphatidic acid is a $P2Y_9R$ agonist)[11]. According to their phylogenetic and sequence divergence, two distinct P2YR subgroups have been proposed. The first group includes the $P2Y_{1/2/4/6/11}R$ subtypes, with a sequence homology of 35–52% in amino acid composition and the presence of a Y-Q/K-X-X-R defining motif in the transmembrane α-helix 7, thus affecting ligand-binding characteristics. The second group contains $P2Y_{12/13/14}R$, with members sharing a sequence homology of 47–48% and the presence of the K-E-X-X-L motif in transmembrane α-helix 7 (ref. 11). There is some evidence suggesting that the two P2YR subgroups also differ in their primary coupling to G proteins: the $P2Y_{1/2/4/6/11}R$ group is coupled to G_q/G_{11} (leading to calcium release via phospholipase C/inositol-1,4,5-triphosphate activation). By contrast, $P2Y_{12/13/14}R$ bind to $G_{i/0}$ proteins, which inhibit adenylate cyclase and modulate flow through ion channels[11,12]. However, there are several instances in which other signalling pathways have been identified. For example, the discrepancy between structural-group affiliation and functional characteristics is highlighted by $P2Y_{12}R$. Despite having only 20–25% sequence homology with $P2Y_{1/2/4/6}R$[12], there is considerable functional similarity. Indeed, $P2Y_{12}$ and $P2Y_2R$ can both activate monomeric G-proteins (such as Rac and/or RhoA)[13], and are the only P2YR subtypes that exhibit agonist-induced desensitization through GPCR kinases[14]. These studies indicate that despite some sequence homology among P2YRs, there are marked differences between individual members of the P2YR family regarding their intracellular signalling cascades.

Endogenous ligands for P2YR

The most abundant and best-characterized endogenous ligand for P2YR is the nucleotide ATP. ATP binds to all P2YRs except $P2Y_6R$ and $P2Y_{14}R$[12]. Its binding characteristics exemplify the complexity of P2YR signalling: at low concentrations it is the only native agonist for $P2Y_{11}R$, but at higher concentrations it functions as a partial agonist for $P2Y_1R$ and $P2Y_{13}R$, or as an antagonist for human $P2Y_4R$ or $P2Y_{12}R$[11,12,15]. Other nucleotides, such as ADP, UTP, UDP or UDP-glucose, exhibit more specificity for individual P2YRs. For example, ADP activates $P2Y_1R$, $P2Y_{12}R$ and $P2Y_{13}R$, whereas UTP primarily binds to $P2Y_2R$ and $P2Y_4R$, and to a lesser extent to $P2Y_6R$, for which UDP is its preferred native ligand. $P2Y_{14}R$ is predominantly

[1]Department of Pneumology, Freiburg University Medical Center, Albert-Ludwigs-University, 79106 Freiburg, Germany. [2]Department of Morphology, Surgery and Experimental Medicine, Section of Pathology, Oncology and Experimental Biology, University of Ferrara, Ferrara, 44121, Italy. [3]Organ Protection Program, Department of Anesthesiology, University of Colorado School of Medicine, Aurora, Colorado 80045, USA.

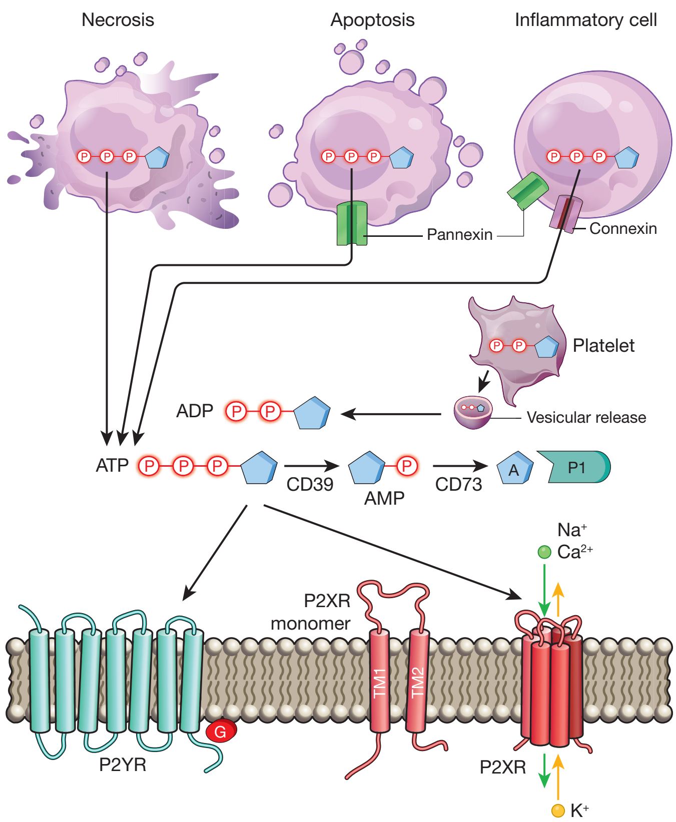


Figure 1 | Extracellular nucleotide release and signalling during inflammation. During inflammation, multiple cell types release nucleotides, for example ATP or ADP, from their intracellular compartments into the extracellular space. Nucleotides can be released during mechanical injury, necrosis, apoptosis or inflammatory cell activation. Several molecular pathways have been implicated in this process, such as vesicular ADP release from platelets, pannexin-mediated ATP release during apoptosis, and connexin- or pannexin-mediated ATP release from inflammatory cells, such as neutrophils. Extracellular nucleotides function as signalling molecules through the activation of purinergic P2 receptors. These receptors can be grouped into metabotropic P2Y receptors (P2YRs; GPCRs with seven transmembrane-spanning motifs) or ionotropic P2X receptors (P2XRs), which are nucleotide-gated ion channels. Each P2XR is formed by three subunits (P2XR monomers), each of which consists of two transmembrane regions, TM1 and TM2. Binding of three molecules of ATP to the assembled P2X channel causes opening of a central pore. These conformational changes allow for flux of ions such as sodium (Na^+), calcium (Ca^{2+}) and potassium (K^+) across the membrane. ATP signalling is terminated by the enzymatic conversion of ATP to adenosine through the ectonucleoside triphosphate diphosphohydrolase CD39 (conversion of ATP/ADP to AMP) and the ecto-5′-nucleotidase CD73 (conversion of AMP to adenosine). Similar to ATP, adenosine (A) functions as an extracellular signalling molecule through the activation of purinergic P1 adenosine receptors.

activated by UDP-glucose and other UDP-sugars, and to a lesser degree by UDP[11,12,15,16]. Indeed, the capacity of different nucleotides to bind specifically to individual P2YRs, or to act as either agonists or antagonists, highlights the complexity of the P2Y system and suggests non-redundant signalling pathways.

Pharmacological compounds that act on P2YR

Owing to the fact that P2YRs have crucial roles in regulating immune responses, they became an obvious pharmacological target for the treatment of inflammatory or infectious diseases. Interestingly the parasiticide suramin, which was widely used in the 1920s for the treatment of human onchocerciasis and trypanosomiasis[17,18], was later found to be a nonspecific inhibitor of P2YR and P2XR[11,15]. Although many P2YR-subtype-specific agonists or antagonists have been characterized in *in vitro* assays or animal studies of inflammatory disorders[11,12,15], presently only two types of P2YR-specific compounds are used in patients: antithrombotic $P2Y_{12}R$ antagonists (for example, clopidrogel) and the $P2Y_2R$ agonist denufosol, which was examined for the treatment of cystic fibrosis, but eventually failed in clinical trials[19]. One of the important future challenges for targeting PYR signalling in patients will include the development of highly selective P2YR antagonists, or specific combined P2R antagonists (for example, a $P2Y_2$/$P2Y_6$/$P2X_7R$ antagonist), which could be used for the treatment of chronic inflammatory disorders.

Functional roles of P2YR in unchallenged mice

Mice with genetic deletions for human P2YR homologous genes have been generated and characterized, with the exception of *P2RY11*, which is not expressed in mice[12]. Despite their widespread expression and their functional involvement in many diseases, mice with global deletions for individual P2YRs display only mild phenotypical alterations when maintained unchallenged in a germ-free environment. For instance, *P2ry2*$^{-/-}$ mice have slightly lower plasma concentrations of aldosterone, renin and potassium[20], whereas global deletion of *P2ry4* is associated with lower exercise capacity and reduced myocardial hypertrophy during a swimming exercise[21]. These findings indicate the likelihood of some redundancy in the signalling system, or compensatory mechanisms following global *P2ry* gene deletion.

P2YR signalling during inflammatory disease states

Several studies over the past decade have highlighted fundamental roles for P2YRs during inflammatory and infectious diseases. Particularly, signalling events through $P2Y_{2/6/12}R$ have shaped an ambivalent view of their function as either friend or foe during inflammation.

$P2Y_2R$

An early attempt at targeting P2Y signalling for the treatment of inflammatory disorders came from studies of $P2Y_2R$ agonists for the treatment of cystic fibrosis[22,23]. Cystic fibrosis is a life-shortening disease that affects over 30,000 children and adults in the United States[24]. The airways of patients with cystic fibrosis are susceptible to infection, characterized by neutrophilic inflammation. Although neutrophil proteases are critical for killing engulfed bacteria, neutrophil elastase accumulates in the airways of cystic fibrosis patients, impairing ciliary function, crippling bacterial clearance and degrading structural proteins[24]. From a molecular perspective, cystic fibrosis is characterized by a defect in the cystic fibrosis transmembrane conductance regulator gene, causing hyperabsorption of sodium leading to thickening of mucus, reduced mucociliary clearance and concomitant increases in susceptibility to bacterial infection[22,23]. Several studies have indicated that $P2Y_2R$ agonists can induce chloride secretion through inhibition of the epithelial sodium channel ENaC, activation of calcium-dependent chloride channels[25], stimulation of mucin production, surfactant secretion and ciliary beating (Fig. 2)[25]. These observations were followed by the development of the $P2Y_2R$ agonist denufosol for the treatment of patients suffering from cystic fibrosis[26]. In 2005 denufosol entered clinical trials, and a 28-day intervention study in a small cohort indicated a potential benefit in lung function of cystic fibrosis patients[27]. Unfortunately, long-term follow-up (48 weeks) in 466 patients was not associated with improved pulmonary function or reduction of pulmonary exacerbations[19]. These disappointing findings may be related to an inflammatory role for $P2Y_2R$ signalling (see below)[2].

In addition to a means for enhancing mucociliary clearance, $P2Y_2R$ agonists have been implicated in the promotion of wound healing[28]. In this context, $P2Y_2R$ signalling mediates the recruitment of leukocytes to the site of tissue damage as well as differentiation and proliferation of structural cells. Moreover, ATP release and concomitant $P2Y_2R$ signalling has been identified as a 'find-me' signal for leukocytes, promoting phagocytic clearance of apoptotic cells or bacteria by macrophages[29] and neutrophils[30,31], thereby contributing to the resolution of inflammation (Fig. 2).

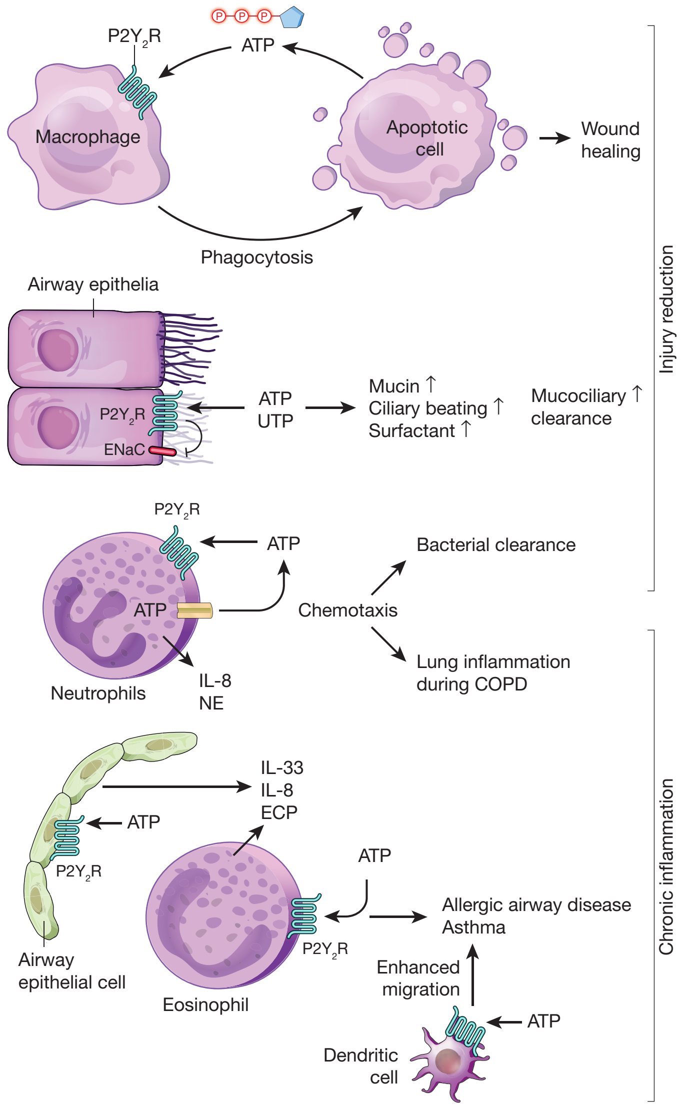


Figure 2 | $P2Y_2R$ signalling during injury resolution and chronic inflammation. $P2Y_2R$ signalling on phagocytes, such as macrophages, contributes to the clearance of apoptotic cells, which release the $P2Y_2R$ agonist ATP as a 'find-me' signal. $P2Y_2R$-mediated clearance of apoptotic cells and debris contributes to wound healing. Activation of $P2Y_2R$ by UTP or ATP promotes mucociliary clearance in the airways via inhibition of the epithelial sodium channel (ENaC), which is associated with concomitant increases in mucin production, surfactant-secretion and ciliary beating. Neutrophil-dependent ATP release and autocrine activation of $P2Y_2R$ contributes to purinergic chemotaxis, thereby enhancing bacterial clearance during pneumonia. On the other hand, $P2Y_2R$-mediated release of IL-8 and neutrophil elastase (NE) from neutrophils contributes to the pathogenesis of chronic obstructive lung disease (COPD). ATP-elicited $P2Y_2R$ signalling on alveolar epithelial cells or eosinophils causes production of pro-allergic mediators (for example, IL-33, IL-8, eosinophil cationic protein) during allergic airway disease. Similarly, $P2Y_2R$ signalling on dendritic cells has a role during the induction and self-perpetuation of asthma.

Other studies have indicated that $P2Y_2R$ signalling contributes to fundamental leukocyte functions such as migration and mediator production by neutrophils, eosinophils, dendritic cells (DCs) and macrophages[32,33]. For example, migrating neutrophils can release ATP from their leading edge to amplify chemotactic signals and direct cell orientation by feedback signalling involving $P2Y_2R$ (Fig. 2)[31]. In addition, danger signals such as uric acid, complement factor 5a, Toll-like receptor ligands and interleukin (IL)-8 are stimulated by an autocrine ATP–$P2Y_2R$ loop to modulate migration and cytokine production of neutrophils or eosinophils[2,31,34–36].

Several studies point towards an ambivalent function of $P2Y_2R$ in neutrophilic inflammation: although *P2ry2*$^{-/-}$ mice are less capable of containing bacterial infections[30,37], inappropriate activation of $P2Y_2R$ is associated with neutrophil-induced hyperinflammation and tissue damage during sepsis, chronic lung disease and hepatitis (Fig. 2)[38–41]. For example, neutrophils from patients with chronic obstructive pulmonary disease express higher levels of $P2Y_2R$, which is associated with higher elastolytic activity and migration capacity upon ATP stimulation compared to healthy controls[40]. In the context of $P2Y_2R$ agonists for the treatment of cystic fibrosis, these findings could explain why the long-term use of inhaled denufosol in cystic fibrosis patients failed to improve clinical outcomes, as it may have been associated with enhanced neutrophil activation and increased lung inflammation, thus overcoming the beneficial effects of improved mucociliary clearance.

Studies in murine models of asthma or contact hypersensitivity demonstrate a contribution of $P2Y_2R$ to the induction and self-perpetuation of allergic diseases. Allergen provocation leads to ATP release and concomitant signalling through $P2Y_2R$, thus favouring the recruitment of immature myeloid DCs and eosinophils to the site of allergen exposures[42–44]. This is associated with the production of pro-allergic mediators (for example, IL-33, IL-8, eosinophil cationic protein) from different cellular sources (Fig. 2)[33,45,46]. Similarly, studies in humans indicate that $P2Y_2R$-induced migration and production of reactive oxygen species are enhanced in immature monocyte-derived DCs and eosinophils from allergic donors in the context of concomitant increases in $P2Y_2R$ expression[43].

Taken together, these findings suggest that ATP-elicited activation of $P2Y_2R$ can function as a 'friend' in the defence against bacterial infections, promotion of wound healing or in enhancement of mucociliary clearance mechanisms. By contrast, it can also lead to uncontrolled inflammation, attenuated resolution, promotion of chronic inflammatory disease states and fibrotic remodelling[47]. Indeed, $P2Y_2R$ antagonists—as opposed to $P2Y_2R$ agonists—could evolve as useful drugs for the treatment for chronic inflammatory diseases, such as chronic obstructive pulmonary disease (Fig. 2)[40,41].

$P2Y_6R$

$P2Y_6R$ is highly expressed on stromal cells and can be activated by UDP[48–50]. The ambivalent behaviour ('friend or foe') of $P2Y_2R$ signalling during inflammatory diseases also applies to $P2Y_6R$ signalling. Recent *in vivo* studies demonstrate a role of $P2Y_6R$ in innate immune responses against bacterial infection[51]. $P2Y_6R$ activation triggers chemokine release from monocytes, DCs, eosinophils and endothelial cells, thus promoting recruitment of inflammatory cells towards the site of inflammation or infection[11,48,51–53]. Similarly, one study demonstrated that injured neurons release UTP and UDP, causing the upregulation of $P2Y_6R$ expression on microglia, and concomitant enhancement of their phagocytic capacity for dying cells[54]. UDP signalling through $P2Y_6R$ can therefore function as an 'eat-me' signal for microglia, thereby initiating the clearance of dying cells or debris in the central nervous system.

By contrast, $P2Y_6R$ signalling is detrimental in models of endothelial[50] or epithelial[49] inflammation. Mucosal $P2Y_6R$ expression is increased during experimentally induced intestinal inflammation such as occurs during inflammatory bowel disease (IBD)[49]. IBD is a heterogeneous group of disorders characterized by intestinal inflammation, including Crohn's disease and ulcerative colitis. Here, pharmacological inhibition or genetic deletion of *P2ry6* in murine models of intestinal inflammation is associated with improved disease outcomes[55]. Similarly, a functional role for $P2Y_6R$ signalling in promoting detrimental inflammation has been reported for chronic forms of lung disease, such as asthma[48]. Indeed, its functional role in promoting pathological airway inflammation during chronic lung disease is an important concern regarding potential considerations for the use of $P2Y_6R$ agonists for the treatment of cystic fibrosis as a means towards enhancing mucociliary clearance[25]. The idea that UDP-elicited $P2Y_6R$ activation can lead to self-perpetuating chronic inflammatory disorders is further supported by recent *in vivo* findings suggesting a functional role of $P2Y_6R$ in promoting atherosclerotic disease in murine models[56]. Taken together, these studies indicate that although activation of $P2Y_6R$ is important

in initiating innate immune responses after infection, inappropriate $P2Y_6R$ signalling, predominantly on stromal cells, can drive detrimental immune responses in chronic inflammatory disorders such as atherosclerosis, chronic lung disease or IBD.

$P2Y_{12}R$

$P2Y_{12}R$, which is highly expressed on platelets, has fundamental roles in platelet activation and aggregation. Stimulation of $P2Y_{12}R$ inhibits adenylyl cyclase activity and increases phosphatidylinositol-3 kinase activity, resulting in the activation of the fibrinogen receptor (integrin αIIb β3), which is critical for platelet aggregation[5]. $P2Y_{12}R$ antagonists have been used successfully for antithrombotic therapy in patients[5]. Because platelets are a key source of inflammatory mediators, $P2Y_{12}R$ signalling has also been implicated in modulating inflammatory responses[57]. The fact that $P2Y_{12}R$ agonists trigger mediator release from platelets implicates inflammatory alternations in patients taking $P2Y_{12}R$ antagonists. Importantly, $P2Y_{12}R$ antagonists such as clopidogrel or ticagrelor are clinically used in patients as platelet inhibitors. Indeed, reduced levels of circulating inflammatory mediators (for example, tumour-necrosis factor-α, C-reactive protein, P-selectin) were found in patients receiving clopidogrel[58]. Preclinical studies confirm the proinflammatory role of $PY_{12}R$ signalling in models of vascular inflammation and asthma. For instance *P2ry12*$^{-/-}$ mice are protected in models of atherosclerosis[59,60]. Moreover, a surprising crosstalk involving leukotrienes and $P2Y_{12}R$ has been described during asthma. In brief, murine studies demonstrate that $P2Y_{12}R$ signalling on platelets is required for the pro-asthmatic action of leukotriene LTE4 (ref. 61). Furthermore, platelet-independent $P2Y_{12}R$ signalling events contribute to asthma, as $P2Y_{12}R$ antagonists can directly block cysteinyl leukotriene-induced release of eosinophil cationic protein from human eosinophils[62], and ADP-elicited $P2Y_{12}R$ activation enhances the capacity of DCs to activate allergen-specific T cells[63]. The clinical observation that single nucleotide polymorphisms of the *P2RY12* gene are associated with altered lung function in a cohort of asthmatic children provides additional evidence for a role of $P2Y_{12}R$ signalling in human asthma[64]. Taken together, these studies implicate $P2Y_{12}R$ signalling in promoting chronic inflammatory disorders such as asthma and atherosclerosis. However, additional clinical trials addressing the clinical efficacy of $P2Y_{12}R$ antagonists (such as clopidogrel or ticagrelor) for the treatment or prevention of chronic inflammation will be critical in establishing their clinical usefulness beyond current indications.

P2XR signalling during inflammation

Molecular structure and signalling cascade of P2XR

P2XRs are plasma membrane channels selective for monovalent and divalent cations (Na^+, K^+, Ca^{2+}) which are directly activated by extracellular ATP (Fig. 1)[65]. Seven different subunits have been identified so far ($P2X_{1-7}R$)[65]. The primary sequence of P2XRs has no important sequence homology with other ligand-gated ion channels, ATP-binding proteins or other known proteins[66]. P2XRs share a common topology with two transmembrane domains (TM1 and TM2), a large extracellular loop responsible for ligand binding, and an intracellular N and a longer C terminus[67]. The extracellular 'loop' starts in proximity of position 52, and ends near proline 329, therefore most of the P2XR protein protrudes from the plasma membrane. Evidence suggests that functional P2XRs are trimers, with three peptide subunits arranged around an ion-permeable channel pore, where ATP binding promotes subunit rearrangement and ion channel opening[68]. Three molecules of ATP seem to bind to the extracellular portions of P2XR[67]. Channel opening induces transmembrane ion fluxes, that is, Na^+ and Ca^{2+} influx and K^+ efflux, leading to plasma membrane depolarization, and—due to the increase of intracellular Ca^{2+} levels—activation of Ca^{2+} signalling cascades, such as p38 MAPK or phospholipase A_2 activation[67]. Interestingly, $P2X_7R$ is capable of activating the NOD-like-receptor-mediated inflammasome assembly with pro-caspase 1 proteolytic activation and subsequent pro-IL-1 and pro-IL-18 cleavage and release of their biologically active forms[65]. The long C-terminal 'tail' of $P2X_7R$ allows it to undergo a conformational change resulting in the so called 'permeability transition'; that is, $P2X_7R$ changes from a cationic channel to a wider pore, allowing transmembrane fluxes of small hydrophilic molecules (including ATP) with a molecular mass of approximately 900 (refs 69, 70).

In addition, some P2XRs, for example $P2X_1R$, can form heterotrimers with $P2X_2R$, $P2X_4R$ and $P2X_5R$ subunits, whereas $P2X_2R$ forms hetero-oligomers with $P2X_3R$[66]. Together these findings indicate that despite the fact that all P2XRs are activated by ATP and share significant sequence homology with each other, highly distinct functions are unique for individual members of the P2XR family.

Endogenous and pharmacological ligands for P2XR

In contrast to P2YR signalling, for which more than one native agonist exists, human P2XRs share ATP as their main endogenous agonist[71]. Owing to the central role of $P2X_7R$ during inflammatory disorders, great efforts have been made to develop selective antagonists. For example, the $P2X_7R$ antagonist AZ9056 was used in patients with rheumatoid arthritis. Although initial studies were promising, Phase IIb/III studies with different $P2X_7R$ antagonists failed to improve long-term clinical outcomes[72]. In addition, there are ongoing Phase II studies with $P2X_7R$ and $P2X_3R$ antagonists in chronic pain and chronic cough (http://www.clinicaltrials.gov/)[72], highlighting the potential for P2XRs as drugable targets in the treatment of inflammatory disorders.

Functional roles of P2XR in unchallenged mice

Physiological roles of P2XRs are indicated by phenotypic manifestations in mice lacking individual P2XR subtypes. Although all single *P2rx* knockout mice are viable and survive to adulthood, some of them reveal unexpected phenotypes, suggesting that functional roles of individual receptor subunits cannot be compensated for by others. For instance, *P2rx1*$^{-/-}$ mice show reduced vas deferens contraction and are infertile[73], whereas *P2rx2*$^{-/-}$ mice develop severe progressive hearing loss[74], and *P2rx3*$^{-/-}$ mice experience urinary bladder hyporeflexia[75]. Mice lacking both *P2rx2* and *P2rx3* have enlarged spleens and increased numbers of immune cells[76]. These findings suggest P2XR-subtype-specific signalling functions under physiological conditions, including immune functions.

$P2X_7R$ signalling during inflammatory disease

Compelling evidence implicates P2XR during inflammation and immune response against microbes[72,78]. Although several other P2XRs are functional during inflammation (for example, $P2X_4R$), $P2X_7R$ in particular has been shown to affect the outcomes of inflammatory or infectious diseases. This may be due to the fact that $P2X_7Rs$ are predominantly expressed on immune cells such as mast cells, macrophages, microglia and DCs (Supplementary Table 1)[2]. Indeed, functional studies implicate $P2X_7R$ in immune responses against bacterial and parasitic infection. For example, activation of $P2X_7R$ is involved in the formation of macrophage multinucleated giant cells, an important step for the control of tuberculosis (Fig. 3)[77]. Other studies implicate $P2X_7R$ signalling in the elimination of intracellular microbes—such as *Mycobacterium tuberculosis*, *Chlamydia psittaci*, *Leishmania amazonensis* or *Toxoplasma gondii*—by either killing of the pathogen or by inducing cell death of infected macrophages (Fig. 3)[78]. Human studies indicate that loss-of-function mutations in the *P2RX7* gene are associated with increased susceptibility to tuberculosis or toxoplasmosis[78]. Owing to its proinflammatory role via activation of the inflammasome, and its direct cytotoxic or pro-apoptotic function, many reports implicate a role for the ATP–$P2X_7R$ axis in tumour suppression. $P2X_7R$ expression is lower in some types of cancer and loss-of-function mutations in the *P2RX7* gene have been linked to the pathogenesis of chronic lymphatic leukaemia[79]. Additional evidence comes from studies demonstrating that loss-of-function mutations of the *P2RX7* gene in patients with breast cancer are associated with an increased risk of progression to metastatic disease states[80]. This study specifically implicated DC–$P2X_7R$ signalling in resistance to chemotherapy (Fig. 3). DCs present antigens from dying cancer cells to prime tumour-specific interferon-γ (IFN-γ)-producing T cells. Dying tumour cells release ATP, which activates $P2X_7R$ expressed on DCs, which in turn causes inflammasome assembly and subsequent secretion of IL-1β (Fig. 3). Accordingly, anticancer chemotherapy was shown to be inefficient against

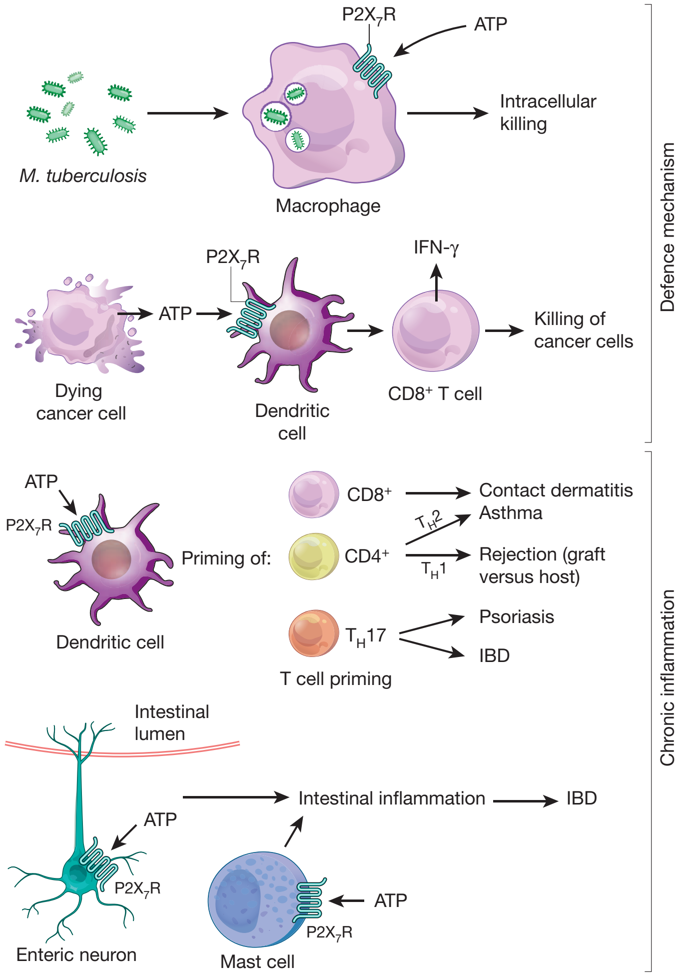


Figure 3 | $P2X_7R$ signalling during infection and inflammation. $P2X_7R$ is required for mounting an appropriate inflammatory response to defend against invading pathogens, for example during intracellular killing of *Mycobacterium tuberculosis* by macrophages. Dying tumour cells release ATP, which activates $P2X_7R$ expressed on DCs, which in turn promotes the priming of IFN-γ-producing cytotoxic $CD8^+$ T cells that kill cancer cells. On the other hand, $P2X_7R$ signalling on DCs and concomitant T-cell priming contributes to allergic disease states, such as $CD8^+$ T-cell-elicited contact dermatitis. DC-mediated T-cell priming under the control of $P2X_7R$ signalling has also been shown to promote T_H1 responses that are implicated in graft-versus-host disease, which contributes to the rejection of a transplanted organ. Similarly, $P2X_7R$-mediated T-cell priming towards a T_H2 response promotes allergic airway disease during asthma. Priming of T_H17 cells is critical during psoriasis and contributes to intestinal inflammation as occurs during IBD. $P2X_7R$ signalling on enteric neurons or mast cells has been implicated in promoting intestinal inflammation during IBD.

tumours established in purinergic receptor *P2rx7*$^{-/-}$ hosts[80]. Together, such findings implicate ATP signalling through the $P2X_7R$ in host-defence mechanisms against intracellular pathogens and cancers.

It is particularly interesting that $P2X_7R$ signalling on DCs can have very different effects on T-cell priming, depending on the specific context, including $CD8^+$ and $CD4^+$ T-cell differentiation (Fig. 3). In contrast to its essential role in immune priming for response against tumours or pathogens, the involvement of $P2X_7R$ in the polarization of antigen-specific effector T cells by DCs contributes to the induction and maintenance of chronic inflammation. $P2X_7R$ signalling on DCs is involved in the sensitization phase of allergic disorders such as contact hypersensitivity ($CD8^+$ T-cell priming)[44] and asthma ($CD4^+$ T-cells, T_H2 response)[81], and contributes to transplant rejection ($CD4^+$ T cells, T_H1 response; Fig. 3). For example, recent studies implicate $P2X_7R$ signalling in graft-versus-host disease, a common complication following an allogeneic tissue transplant, in which immune cells in the tissue recognize the recipient as 'foreign', leading to an immunological reaction of transplanted immune cells against the host. Indeed, $P2X_7R$ inhibition or deficiency on DCs is associated with reduced severity of graft-versus-host disease[82]. As such, $P2X_7R$ signalling of antigen-presenting DCs led to an increased expression of CD80 and CD86 *in vitro* and *in vivo* and activated a cascade of proinflammatory events, including signal transducer and activator of transcription 1 (STAT1) phosphorylation, IFN-γ production and donor T-cell expansion[82]. Again, other studies report that $P2X_7R$-mediated priming can contribute to T_H17-driven autoimmune diseases, such as psoriasis (Fig. 3)[83]. By triggering the production of pro-allergic mediators from eosinophils, mast cells, macrophages and DCs, $P2X_7R$ signalling also contributes to the effector phase and chronification of allergic disorders[42,81]. Furthermore, increased expression of $P2X_7R$ can be found on eosinophils and macrophages in asthma (Fig. 3)[81], and loss-of-function mutations in the *P2rx7* gene have been associated with attenuated risk of allergen sensitization and asthma[84].

In line with these findings, other studies report a detrimental role of $P2X_7R$ in promoting excessive inflammation during IBD, by showing that ATP derived from commensal bacteria activates a unique subset of lamina propria cells ($CD70^{high}$ $CD11c^{low}$ cells), leading to T_H17 cell differentiation (Fig. 3)[85]. Indeed, germ-free mice exhibit lower concentrations of luminal ATP, accompanied by fewer lamina propria T_H17 cells[85]. $P2X_7R$ also participates in IBD pathogenesis by mediating enteric neural death (Fig. 3)[85,86]. Finally, mast-cell-dependent mechanisms of intestinal inflammation are under the control of $P2X_7R$, as increased $P2X_7R$ expression can be found in mast cells from Crohn's patients and inhibition of $P2X_7R$ on mast cells dampens intestinal inflammation (Fig. 3)[87].

Together, these findings expose $P2X_7R$ signalling during inflammation as a double-edged sword: these receptors have a critical role in mediating appropriate inflammatory and immunological responses against invading pathogens or cancer cells, respectively, but contribute to chronic inflammatory disease states in a wide range of inflammatory disorders, such as chronic lung disease[40,88], asthma[81,84] or IBD[86,87], when activated inappropriately.

Termination of ATP signalling

Termination of P2R signalling involves the conversion of ATP/ADP to adenosine within the extracellular compartment by the activity of ecto-nucleotidases. The four main groups of ectonucleotidases are the ectonucleoside triphosphate diphosphohydrolases (NTPDases), ecto-5′-nucleotidase (CD73), ectonucleotide pyrophosphatase/phosphodiesterases and alkaline phosphatases[89]. NTPDases represent a family of ubiquitously expressed membrane-bound enzymes. The catalytic sites of plasma membrane-expressed NTPDases 1–3 and 8 are oriented towards the extracellular milieu[90]. Owing to its high expression in many tissues and its ability to catalyse the conversion of ATP (and ADP) down to AMP, many studies have found a functional role for NTPDase1 (CD39) in the termination of P2R signalling[90,91]. Next, extracellular AMP is converted to adenosine by CD73 (Fig. 1)[5]. Therefore, termination of ATP signalling is closely linked to the generation of extracellular adenosine. In many instances, adenosine-elicited P1R signalling dampens acute inflammation and tissue injury[92,93], thus opposing inflammatory functions of P2Rs (Fig. 1)[94,95].

Consistent with a protective role for the CD39/CD73 pathway in terminating inflammatory P2R signalling, and concomitantly increasing extracellular adenosine levels and signalling events, several studies show that *Cd39*$^{-/-}$ or *Cd73*$^{-/-}$ mice are prone to tissue injury during inflammatory conditions such as acute lung injury or intestinal inflammation[96,97]. For example, patients with a single nucleotide polymorphism associated with low levels of CD39 expression have increased susceptibility to Crohn's disease, suggesting that deficiency in CD39 could be associated with IBD in humans[96]. Other reports suggest that CD39 exerts a protective thromboregulatory function in stroke by preventing P2R-mediated thrombosis[91]. Moreover, several studies implicate the CD39/CD73 pathway in the immunosuppressive roles of regulatory T cells (T_{reg}). These are a group of $CD4^+$ lymphocytes that suppress T-cell responses against a variety of pathogens and control inappropriate immune activation, thus limiting collateral tissue

damage but allowing pathogen persistence[98]. As such, T_{reg} cells from *Cd39*$^{-/-}$ mice demonstrate attenuated suppressive functions *in vitro* and fail to block rejection of allografts *in vivo*[99]. Similarly, *Cd73*$^{-/-}$ mice fail to resolve lung injury induced by lipopolysaccharide inhalation due to impairment of T_{reg} functions[100].

The ambivalence of CD39/CD73-mediated control of T_{regs} is further exemplified during infections with human immunodeficiency virus (HIV), the retrovirus known to cause AIDS in humans. HIV infections are characterized by a progressive CD4 lymphopenia in conjunction with defective HIV-specific CD8 responses that are critical for the control of viral replication[98]. As such, the consequences of T_{reg} expansion, as seen during HIV infection, could have either a beneficial function by suppressing generalized T-cell activation, or could be fatal owing to attenuated HIV-specific responses and thus promoting viral persistence[98]. For example, studies demonstrate that HIV-1-positive patients have an increase of T_{reg}-associated expression of CD39. These findings indicate that the CD39/CD73 pathway is involved in T_{reg} suppression in HIV infection[101]. A genetic association study demonstrated that a polymorphism in the *CD39* gene is associated with attenuated CD39 expression and slower progression to AIDS in HIV-infected patients[101]. Thus, it can be speculated that CD39$^+$ T_{regs} are the most potent T_{reg} subset to inhibit HIV-specific T-cell responses. This could at least in part account for their association with disease progression. Other examples for a detrimental role of CD39-dependent ATP breakdown come from studies of autoimmune hepatitis in which natural killer T cell dysfunction in *Cd39*$^{-/-}$ mice protects against concanavalin A-induced hepatitis. Heightened levels of apoptosis of *Cd39*$^{-/-}$ natural killer T cells *in vivo* and *in vitro* appear to be driven by unimpeded activation of $P2X_7R$[102]. Similarly, enzymatic removal of ATP by apyrase (conversion of ATP/ADP to AMP) or ectopic CD39 expression attenuates clearance of apoptotic cells, indicating a detrimental role for CD39-dependent ATP phosphohydrolysis in dampening efficient corpse clearance via $P2Y_2R$ signalling[29]. Other studies demonstrate the existence of bacterial ectotriphosphate diphosphohydrolases—similar to human CD39—which are critical for the intracellular multiplication of *Legionella pneumophila* by preventing P2R-elicited immune responses. As such, these findings implicate bacterial ectotriphosphate diphosphohydrolases in virulence[103]. Together these studies exemplify an ambivalent role for the termination of P2R signalling via enzymatic phosphohydrolysis. Although this pathway is critical in preventing excessive P2R-dependent inflammation in a sterile environment[104], CD39 function can become detrimental for the appropriate clearance of apoptotic debris[29], inflammation directed against bacterial infections[103] or by generating an immunosuppressive environment, which promotes the development or progression of cancer[5].

Functional role of P1 signalling during inflammation

Extracellular AMP generated by phosphohydrolysis of precursor nucleotides (for example, ATP or ADP) has no clearly characterized signalling function (for example, through specific AMP receptors). However, extracellular AMP serves as the metabolic substrate for the extracellular generation of adenosine via CD73 (Fig. 1)[5]. Once generated within the extracellular compartment, adenosine can function via activation of four distinct P1 receptors: ADORA1, ADORA2A, ADORA2B or ADORA3. Adenosine signalling is terminated via uptake of adenosine from the extracellular towards the intracellular compartment through equilibrative nucleoside transporters and is metabolized to inosine via adenosine deaminase[105], or to AMP via adenosine kinase[106]. Several studies implicate adenosine signalling in dampening excessive inflammation[92]. For example, *Adora2a*$^{-/-}$ mice experience increased inflammation including extensive tissue damage, more prolonged and higher levels of proinflammatory cytokines, and mortality when exposed to sub-threshold doses of inflammatory stimuli[93]. Other studies demonstrate that ADORA2B signalling dampens excessive inflammation during acute lung injury[107], promotes ischaemia tolerance and improves anaerobic carbohydrate metabolism[108,109]. Similarly, genetic deletion or pharmacological blockade of equilibrative nucleoside transporters is associated with increased adenosine levels and improved outcomes during inflammatory disease states[110,111]. In most instances, the anti-inflammatory signalling effects of adenosine are associated with improved outcomes during inflammatory diseases such as IBD[112], or during sepsis induced by caecal ligation and puncture[113]. However, other studies indicate that the anti-inflammatory effects of adenosine signalling can be detrimental in containing an infection with live bacteria. For example, a recent study demonstrates that antagonism of P1 receptors (for example, ADORA2B) can be useful in enhancing macrophage-mediated bacterial phagocytosis and improving polymicrobial sepsis survival in mice[114]. Together, these studies highlight that P1 and P2 receptors frequently have opposing effects in biological systems, and that shifting the balance from purinergic P2YR and P2XR signalling towards adenosine-mediated P1 signalling is an important therapeutic concept in efforts to dampen pathological inflammation and promote healing[5].

Conclusions

The field of extracellular nucleotide signalling and metabolism is a dynamic area of research with important opportunities for novel treatments for inflammatory or infectious diseases. On the one hand, P2R signalling functions to coordinate appropriate immune responses against invading pathogens or tumours. Indeed, pharmacological approaches that amplify extracellular ATP signalling hold promise as therapies for the treatment of cancer or during uncontrolled infections with live pathogens. Such strategies could include inhibition of ATP breakdown (for example, via nucleotidase inhibitors) or treatment with P2 receptor agonists. Conversely, inadequate P2R signalling has been associated with excessive inflammation, chronification and inappropriate resolution and fibrosis in a wide range of inflammatory diseases. In this context, treatment strategies that block P2R signalling, promote extracellular conversion of ATP to adenosine and activate adenosine receptors have been implicated in the treatment of acute or chronic inflammatory diseases. We therefore anticipate that compounds targeting these pathways will be further exploited in the treatment of inflammatory conditions in human patients in the near future.

Note added in proof: Two reports appeared online regarding the atomic structure of the $P2Y_{12}R$ while the current review was in press. The first report provides a 2.6 Å resolution crystal structure of the human $P2Y_{12}R$ in complex with the non-nucleotide antagonist AZD1283 (ref. 115), thus providing important insights for the development of $P2Y_{12}R$ ligands and allosteric modulators as drug candidates. The second report provides the structures of the human $P2Y_{12}R$ in complex with a full agonist (2-methylthio-adenosine-5′-diphosphate) at a resolution of 2.5 Å, and the corresponding ATP derivative 2-methylthio-adenosine-5′-triphosphate at 3.1 Å resolution[116]. The agonist-bound P2Y12R structure answers ambiguities surrounding P2Y12R-agonist recognition, and suggests unexpected interactions with several residues.

Received 28 August 2013; accepted 28 January 2014.

1. Khakh, B. S. & Burnstock, G. The double life of ATP. *Sci. Am.* **301,** 84–92 (2009).
2. Junger, W. G. Immune cell regulation by autocrine purinergic signalling. *Nature Rev. Immunol.* **11,** 201–212 (2011).
Provides a comprehensive overview of the functional roles of purinergic signalling events on cells of the adaptive and innate immune systems.
3. Fredholm, B. & Verkhratsky, A. Purines — 80 years and very much alive. *Acta Physiol.* **199,** 91–92 (2010).
4. Burnstock, G. Purinergic signalling and disorders of the central nervous system. *Nature Rev. Drug. Discov.* **7,** 575–590 (2008).
5. Eltzschig, H. K., Sitkovsky, M. V. & Robson, S. C. Purinergic signaling during inflammation. *N. Engl. J. Med.* **367,** 2322–2333 (2012).
6. Eltzschig, H. K., Macmanus, C. F. & Colgan, S. P. Neutrophils as sources of extracellular nucleotides: functional consequences at the vascular interface. *Trends Cardiovasc. Med.* **18,** 103–107 (2008).
7. Eltzschig, H. K. *et al.* ATP release from activated neutrophils occurs via connexin 43 and modulates adenosine-dependent endothelial cell function. *Circ. Res.* **99,** 1100–1108 (2006).
8. Chekeni, F. B. *et al.* Pannexin 1 channels mediate 'find-me' signal release and membrane permeability during apoptosis. *Nature* **467,** 863–867 (2010).
Identifies a functional role for pannexin-mediated ATP release from cells undergoing apoptosis as a 'find-me' signal for phagocytes.
9. Faigle, M., Seessle, J., Zug, S., El Kasmi, K. C. & Eltzschig, H. K. ATP release from vascular endothelia occurs across Cx43 hemichannels and is attenuated during hypoxia. *PLoS ONE* **3,** e2801 (2008).

10. Lazarowski, E. R. Vesicular and conductive mechanisms of nucleotide release. *Purinergic Signal.* **8,** 359–373 (2012).
11. Abbracchio, M. P. *et al.* International Union of Pharmacology LVIII: update on the P2Y G protein-coupled nucleotide receptors: from molecular mechanisms and pathophysiology to therapy. *Pharmacol. Rev.* **58,** 281–341 (2006).
12. Jacobson, K. A., Balasubramanian, R., Deflorian, F. & Gao, Z. G. G protein-coupled adenosine (P1) and P2Y receptors: ligand design and receptor interactions. *Purinergic Signal.* **8,** 419–436 (2012).
Provides an update on the medicinal chemistry and pharmacology of the different subtypes of adenosine receptors and P2Y receptors, including recent advances in the identification and characterization of selective ligands.
13. Soulet, C. *et al.* G_i-dependent and -independent mechanisms downstream of the P2Y12 ADP-receptor. *J. Thromb. Haemost.* **2,** 135–146 (2004).
14. Hardy, A. R. *et al.* P2Y1 and P2Y12 receptors for ADP desensitize by distinct kinase-dependent mechanisms. *Blood* **105,** 3552–3560 (2005).
15. von Kugelgen, I. Pharmacological profiles of cloned mammalian P2Y-receptor subtypes. *Pharmacol. Ther.* **110,** 415–432 (2006).
16. Harden, T. K., Sesma, J. I., Fricks, I. P. & Lazarowski, E. R. Signalling and pharmacological properties of the P2Y receptor. *Acta Physiol.* **199,** 149–160 (2010).
17. Hawking, F. Suramin: with special reference to onchocerciasis. *Adv. Pharmacol. Chemother.* **15,** 289–322 (1978).
18. Voogd, T. E., Vansterkenburg, E. L., Wilting, J. & Janssen, L. H. Recent research on the biological activity of suramin. *Pharmacol. Rev.* **45,** 177–203 (1993).
19. Ratjen, F. *et al.* Long term effects of denufosol tetrasodium in patients with cystic fibrosis. *J. Cyst. Fibros.* **11,** 539–549 (2012).
20. Rieg, T. *et al.* Mice lacking P2Y2 receptors have salt-resistant hypertension and facilitated renal Na^+ and water reabsorption. *FASEB J.* **21,** 3717–3726 (2007).
21. Horckmans, M. *et al.* Gene deletion of $P2Y_4$ receptor lowers exercise capacity and reduces myocardial hypertrophy with swimming exercise. *Am. J. Physiol. Heart. Circ. Physiol.* **303,** H835–H843 (2012).
22. Knowles, M. R., Clarke, L. L. & Boucher, R. C. Activation by extracellular nucleotides of chloride secretion in the airway epithelia of patients with cystic fibrosis. *N. Engl. J. Med.* **325,** 533–538 (1991).
23. Parr, C. E. *et al.* Cloning and expression of a human P2U nucleotide receptor, a target for cystic fibrosis pharmacotherapy. *Proc. Natl Acad. Sci. USA* **91,** 3275–3279 (1994).
24. Davis, S. D. & Ferkol, T. Identifying the origins of cystic fibrosis lung disease. *N. Engl. J. Med.* **368,** 2026–2028 (2013).
25. Burnstock, G., Brouns, I., Adriaensen, D. & Timmermans, J. P. Purinergic signaling in the airways. *Pharmacol. Rev.* **64,** 834–868 (2012).
26. Kellerman, D. *et al.* Denufosol: a review of studies with inhaled $P2Y_2$ agonists that led to Phase 3. *Pulm. Pharmacol. Ther.* **21,** 600–607 (2008).
27. Deterding, R. R. *et al.* Phase 2 randomized safety and efficacy trial of nebulized denufosol tetrasodium in cystic fibrosis. *Am. J. Respir. Crit. Care Med.* **176,** 362–369 (2007).
28. Gendaszewska-Darmach, E. & Kucharska, M. Nucleotide receptors as targets in the pharmacological enhancement of dermal wound healing. *Purinergic Signal.* **7,** 193–206 (2011).
29. Elliott, M. R. *et al.* Nucleotides released by apoptotic cells act as a find-me signal to promote phagocytic clearance. *Nature* **461,** 282–286 (2009).
30. Chen, Y. *et al.* Purinergic signaling: a fundamental mechanism in neutrophil activation. *Sci. Signal* **3,** ra45 (2010).
31. Chen, Y. *et al.* ATP release guides neutrophil chemotaxis via P2Y2 and A3 receptors. *Science* **314,** 1792–1795 (2006).
Shows that human neutrophils release ATP from the leading edge of the cell surface to amplify chemotactic signals and direct cell orientation by feedback through $P2Y_2R$ as a mechanism of purinergic chemotaxis.
32. Myrtek, D. & Idzko, M. Chemotactic activity of extracellular nucleotideson human immune cells. *Purinergic Signal.* **3,** 5–11 (2007).
33. Ferrari, D. *et al.* Activation of human eosinophils via P2 receptors: novel findings and future perspectives. *J. Leukoc. Biol.* **79,** 7–15 (2006).
34. Kronlage, M. *et al.* Autocrine purinergic receptor signaling is essential for macrophage chemotaxis. *Sci. Signal* **3,** ra55 (2010).
35. Ben Yebdri, F., Kukulski, F., Tremblay, A. & Sevigny, J. Concomitant activation of $P2Y_2$ and $P2Y_6$ receptors on monocytes is required for TLR1/2-induced neutrophil migration by regulating IL-8 secretion. *Eur. J. Immunol.* **39,** 2885–2894 (2009).
36. Kobayashi, T., Kouzaki, H. & Kita, H. Human eosinophils recognize endogenous danger signal crystalline uric acid and produce proinflammatory cytokines mediated by autocrine ATP. *J. Immunol.* **184,** 6350–6358 (2010).
37. Geary, C. *et al.* Increased susceptibility of purinergic receptor-deficient mice to lung infection with *Pseudomonas aeruginosa*. *Am. J. Physiol. Lung Cell Mol. Physiol.* **289,** L890–L895 (2005).
38. Inoue, Y., Chen, Y., Hirsh, M. I., Yip, L. & Junger, W. G. A3 and P2Y2 receptors control the recruitment of neutrophils to the lungs in a mouse model of sepsis. *Shock* **30,** 173–177 (2008).
39. Ayata, C. K. *et al.* Purinergic $P2Y_2$ receptors promote neutrophil infiltration and hepatocyte death in mice with acute liver injury. *Gastroenterology* **143,** 1620–1629 (2012).
40. Lommatzsch, M. *et al.* Extracellular adenosine triphosphate and chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* **181,** 928–934 (2010).
41. Cicko, S. *et al.* Purinergic receptor inhibition prevents the development of smoke-induced lung injury and emphysema. *J. Immunol.* **185,** 688–697 (2010).
42. Idzko, M. *et al.* Extracellular ATP triggers and maintains asthmatic airway inflammation by activating dendritic cells. *Nature Med.* **13,** 913–919 (2007).
Shows that allergen challenge causes acute accumulation of ATP in the airways of asthmatic subjects or in mice with experimentally induced asthma, and further implicates purinergic signalling as a key mediator in allergen-driven lung inflammation.
43. Muller, T. *et al.* The purinergic receptor $P2Y_2$ receptor mediates chemotaxis of dendritic cells and eosinophils in allergic lung inflammation. *Allergy* **65,** 1545–1553 (2010).
44. Weber, F. C. *et al.* Lack of the purinergic receptor $P2X_7$ results in resistance to contact hypersensitivity. *J. Exp. Med.* **207,** 2609–2619 (2010).
Identifies $P2X_7R$ as a crucial receptor for extracellular ATP released in skin in response to contact allergens, and triggering of contact hypersensitivity.
45. Idzko, M. *et al.* Stimulation of P2 purinergic receptors induces the release of eosinophil cationic protein and interleukin-8 from human eosinophils. *Br. J. Pharmacol.* **138,** 1244–1250 (2003).
46. Kouzaki, H., Iijima, K., Kobayashi, T., O'Grady, S. M. & Kita, H. The danger signal, extracellular ATP, is a sensor for an airborne allergen and triggers IL-33 release and innate Th2-type responses. *J. Immunol.* **186,** 4375–4387 (2011).
47. Kunzli, B. M. *et al.* Upregulation of CD39/NTPDases and P2 receptors in human pancreatic disease. *Am. J. Physiol. Gastrointest. Liver Physiol.* **292,** G223–G230 (2007).
48. Vieira, R. P. *et al.* Purinergic receptor type 6 contributes to airway inflammation and remodeling in experimental allergic airway inflammation. *Am. J. Respir. Crit. Care Med.* **184,** 215–223 (2011).
49. Grbic, D. M., Degagne, E., Langlois, C., Dupuis, A. A. & Gendron, F. P. Intestinal inflammation increases the expression of the $P2Y_6$ receptor on epithelial cells and the release of CXC chemokine ligand 8 by UDP. *J. Immunol.* **180,** 2659–2668 (2008).
50. Riegel, A. K. *et al.* Selective induction of endothelial $P2Y_6$ nucleotide receptor promotes vascular inflammation. *Blood* **117,** 2548–2555 (2011).
51. Zhang, Z. *et al.* $P2Y_6$ agonist uridine 5′-diphosphate promotes host defense against bacterial infection via monocyte chemoattractant protein-1-mediated monocytes/macrophages recruitment. *J. Immunol.* **186,** 5376–5387 (2011).
52. Idzko, M. *et al.* Characterization of the biological activities of uridine diphosphate in human dendritic cells: influence on chemotaxis and CXCL8 release. *J. Cell. Physiol.* **201,** 286–293 (2004).
53. Ferrari, D. *et al.* P2 purinergic receptors of human eosinophils: characterization and coupling to oxygen radical production. *FEBS Lett.* **486,** 217–224 (2000).
54. Koizumi, S. *et al.* UDP acting at $P2Y_6$ receptors is a mediator of microglial phagocytosis. *Nature* **446,** 1091–1095 (2007).
Demonstrates that $P2Y_6R$ is upregulated when neurons are damaged, and implicates its signalling function as a sensor for phagocytosis by sensing diffusible UDP signals, a novel pathway mediating microglial phagocytosis.
55. Grbic, D. M. *et al.* P2Y6 receptor contributes to neutrophil recruitment to inflamed intestinal mucosa by increasing CXC chemokine ligand 8 expression in an AP-1-dependent manner in epithelial cells. *Inflamm. Bowel Dis.* **18,** 1456–1469 (2012).
56. Guns, P. J., Hendrickx, J., Van Assche, T., Fransen, P. & Bult, H. P2Y receptors and atherosclerosis in apolipoprotein E-deficient mice. *Br. J. Pharmacol.* **159,** 326–336 (2010).
57. Semple, J. W., Italiano, J. E., Jr. & Freedman, J. Platelets and the immune continuum. *Nature Rev. Immunol.* **11,** 264–274 (2011).
58. Muhlestein, J. B. Effect of antiplatelet therapy on inflammatory markers in atherothrombotic patients. *Thromb. Haemost.* **103,** 71–82 (2010).
59. Li, D. *et al.* Roles of purinergic receptor P2Y, G protein-coupled 12 in the development of atherosclerosis in apolipoprotein E-deficient mice. *Arterioscler. Thromb. Vasc. Biol.* **32,** e81–e89 (2012).
60. Yashiro, K. *et al.* Involvement of platelet activation by P2Y12 receptor in the development of transplant arteriosclerosis in mice. *Transplantation* **87,** 660–667 (2009).
61. Paruchuri, S. *et al.* Leukotriene E4-induced pulmonary inflammation is mediated by the P2Y12 receptor. *J. Exp. Med.* **206,** 2543–2555 (2009).
62. Neves, J. S., Radke, A. L. & Weller, P. F. Cysteinyl leukotrienes acting via granule membrane-expressed receptors elicit secretion from within cell-free human eosinophil granules. *J. Allergy Clin. Immunol.* **125,** 477–482 (2010).
63. Ben Addi, A., Cammarata, D., Conley, P. B., Boeynaems, J. M. & Robaye, B. Role of the P2Y12 receptor in the modulation of murine dendritic cell function by ADP. *J. Immunol.* **185,** 5900–5906 (2010).
64. Bunyavanich, S., Boyce, J. A., Raby, B. A. & Weiss, S. T. Gene-by-environment effect of house dust mite on purinergic receptor P2Y12 (*P2RY12*) and lung function in children with asthma. *Clin. Exp. Allergy* **42,** 229–237 (2012).
65. Surprenant, A. & North, R. A. Signaling at purinergic P2X receptors. *Annu. Rev. Physiol.* **71,** 333–359 (2009).
66. Khakh, B. S. & North, R. A. Neuromodulation by extracellular ATP and P2X receptors in the CNS. *Neuron* **76,** 51–69 (2012).
67. Khakh, B. S. & North, R. A. P2X receptors as cell-surface ATP sensors in health and disease. *Nature* **442,** 527–532 (2006).
Describes P2X signalling beyond its function in the autonomic nervous system, with particular focus on its key roles in afferent signalling, chronic pain and autocrine loops of endothelial and epithelial cells.
68. Kawate, T., Michel, J. C., Birdsong, W. T. & Gouaux, E. Crystal structure of the ATP-gated $P2X_4$ ion channel in the closed state. *Nature* **460,** 592–598 (2009).
Presents the crystal structure of zebrafish $P2X_4R$ in its closed, resting state, including definition of the locations of three non-canonical, intersubunit ATP-binding sites, and evidence suggesting that ATP binding promotes subunit rearrangement and ion channel opening.

69. Alberto, A. V. *et al.* Is pannexin the pore associated with the P2X7 receptor? *Naunyn Schmiedebergs Arch. Pharmacol.* **386,** 775–787 (2013).
70. Surprenant, A., Rassendren, F., Kawashima, E., North, R. A. & Buell, G. The cytolytic P_{2Z} receptor for extracellular ATP identified as a $P2_X$ receptor ($P2X_7$). *Science* **272,** 735–738 (1996).
71. Schwarz, N. *et al.* Alternative splicing of the N-terminal cytosolic and transmembrane domains of P2X7 controls gating of the ion channel by ADP-ribosylation. *PLoS One* **7,** e41269 (2012).
72. North, R. A. & Jarvis, M. F. P2X receptors as drug targets. *Mol. Pharmacol.* **83,** 759–769 (2013).
73. Mulryan, K. *et al.* Reduced vas deferens contraction and male infertility in mice lacking P2X1 receptors. *Nature* **403,** 86–89 (2000).
74. Yan, D. *et al.* Mutation of the ATP-gated $P2X_2$ receptor leads to progressive hearing loss and increased susceptibility to noise. *Proc. Natl Acad. Sci. USA* **110,** 2228–2233 (2013).
75. Cockayne, D. A. *et al.* Urinary bladder hyporeflexia and reduced pain-related behaviour in $P2X_3$-deficient mice. *Nature* **407,** 1011–1015 (2000).
76. Coutinho-Silva, R., Knight, G. E. & Burnstock, G. Impairment of the splenic immune system in $P2X_2$/$P2X_3$ knockout mice. *Immunobiology* **209,** 661–668 (2005).
77. Falzoni, S. *et al.* The purinergic P2Z receptor of human macrophage cells. Characterization and possible physiological role. *J. Clin. Invest.* **95,** 1207–1216 (1995).
78. Coutinho-Silva, R. & Ojcius, D. M. Role of extracellular nucleotides in the immune response against intracellular bacteria and protozoan parasites. *Microbes Infect.* **14,** 1271–1277 (2012).
79. Stagg, J. & Smyth, M. J. Extracellular adenosine triphosphate and adenosine in cancer. *Oncogene* **29,** 5346–5358 (2010).
80. Ghiringhelli, F. *et al.* Activation of the NLRP3 inflammasome in dendritic cells induces IL-1β-dependent adaptive immunity against tumors. *Nature Med.* **15,** 1170–1178 (2009).
81. Muller, T. *et al.* A potential role for P2X7R in allergic airway inflammation in mice and humans. *Am. J. Respir. Cell Mol. Biol.* **44,** 456–464 (2011).
82. Wilhelm, K. *et al.* Graft-versus-host disease is enhanced by extracellular ATP activating P2X7R. *Nature Med.* **16,** 1434–1438 (2010).
Reveals the relevance of ATP-induced activation of $P2X_7R$ for graft-versus-host disease development, indicating that the physiological metabolite ATP must be recognized as an endogenous danger signal that has detrimental effects when released into the extracellular space after tissue damage through the activation of recipient antigen-presenting cells.
83. Killeen, M. E., Ferris, L., Kupetsky, E. A., Falo, L., Jr. & Mathers, A. R. Signaling through purinergic receptors for ATP induces human cutaneous innate and adaptive Th17 responses: implications in the pathogenesis of psoriasis. *J. Immunol.* **190,** 4324–4336 (2013).
84. Manthei, D. M. *et al.* Protection from asthma in a high-risk birth cohort by attenuated $P2X_7$ function. *J. Allergy Clin. Immunol.* **130,** 496–502 (2012).
85. Atarashi, K. *et al.* ATP drives lamina propria T_H17 cell differentiation. *Nature* **455,** 808–812 (2008).
86. Gulbransen, B. D. *et al.* Activation of neuronal P2X7 receptor-pannexin-1 mediates death of enteric neurons during colitis. *Nature Med.* **18,** 600–604 (2012).
Provides evidence that activation of neuronal $P2X_7R$ through pannexin 1 underlies neuron death and the subsequent development of abnormal gut motility in IBD, suggesting that this pathway could be targeted to ameliorate the progression of gut dysmotility during intestinal inflammation.
87. Kurashima, Y. *et al.* Extracellular ATP mediates mast cell-dependent intestinal inflammation through P2X7 purinoceptors. *Nature Commun.* **3,** 1034 (2012).
88. Lucattelli, M. *et al.* P2X7 receptor signalling in the pathogenesis of smoke-induced lung inflammation and emphysema. *Am. J. Respir. Cell Mol. Biol.* **44,** 423–429 (2010).
89. Zimmermann, H., Zebisch, M. & Strater, N. Cellular function and molecular structure of ecto-nucleotidases. *Purinergic Signal.* **8,** 437–502 (2012).
Summarizes what is known about the cellular and molecular functions of ATP-hydrolysing ectonucleotidases.
90. Kohler, D. *et al.* CD39/ectonucleoside triphosphate diphosphohydrolase 1 provides myocardial protection during cardiac ischemia/reperfusion injury. *Circulation* **116,** 1784–1794 (2007).
91. Pinsky, D. J. *et al.* Elucidation of the thromboregulatory role of CD39/ectoapyrase in the ischemic brain. *J. Clin. Invest.* **109,** 1031–1040 (2002).
92. Flogel, U. *et al.* Selective activation of adenosine A2A receptors on immune cells by a CD73-dependent prodrug suppresses joint inflammation in experimental rheumatoid arthritis. *Sci. Transl. Med.* **4,** 146ra108 (2012).
Here the authors developed a selective ADORA2A agonist that requires the presence of CD73 to become activated. This compound suppresses joint inflammation in experimental rheumatoid arthritis, while avoiding ADORA2A-mediated vasodilation.
93. Ohta, A. & Sitkovsky, M. Role of G-protein-coupled adenosine receptors in downregulation of inflammation and protection from tissue damage. *Nature* **414,** 916–920 (2001).
First study to provide genetic evidence for a non-redundant role for the P1 adenosine receptor ADORA2A in the physiological negative feedback mechanism for limitation and termination of both tissue-specific and systemic inflammatory responses.
94. Sitkovsky, M. V. *et al.* Physiological control of immune response and inflammatory tissue damage by hypoxia-inducible factors and adenosine A2A receptors. *Annu. Rev. Immunol.* **22,** 657–682 (2004).
95. Colgan, S. P. & Eltzschig, H. K. Adenosine and hypoxia-inducible factor signaling in intestinal injury and recovery. *Annu. Rev. Physiol.* **74,** 153–175 (2012).
96. Friedman, D. J. *et al.* CD39 deletion exacerbates experimental murine colitis and human polymorphisms increase susceptibility to inflammatory bowel disease. *Proc. Natl Acad. Sci. USA* **106,** 16788–16793 (2009).
Provides evidence that CD39 deficiency exacerbates murine colitis and suggests that CD39 polymorphisms are associated with IBD in humans.
97. Eckle, T. *et al.* Identification of ectonucleotidases CD39 and CD73 in innate protection during acute lung injury. *J. Immunol.* **178,** 8127–8137 (2007).
98. Chevalier, M. F. & Weiss, L. The split personality of regulatory T cells in HIV infection. *Blood* **121,** 29–37 (2013).
99. Deaglio, S. *et al.* Adenosine generation catalyzed by CD39 and CD73 expressed on regulatory T cells mediates immune suppression. *J. Exp. Med.* **204,** 1257–1265 (2007).
Identifies CD39 and CD73 as surface markers of T_{regs} and implicates extracellular adenosine generation in an autocrine signalling loop critical for the suppressor functions of T_{regs}.
100. Ehrentraut, H. *et al.* $CD73^+$ regulatory T cells contribute to adenosine-mediated resolution of acute lung injury. *FASEB J.* **27,** 2207–2219 (2013).
101. Nikolova, M. *et al.* CD39/adenosine pathway is involved in AIDS progression. *PLoS Pathog.* **7,** e1002110 (2011).
102. Beldi, G. *et al.* Natural killer T cell dysfunction in CD39-null mice protects against concanavalin A-induced hepatitis. *Hepatology* **48,** 841–852 (2008).
103. Sansom, F. M. *et al.* A bacterial ecto-triphosphate diphosphohydrolase similar to human CD39 is essential for intracellular multiplication of *Legionella pneumophila*. *Cell Microbiol.* **9,** 1922–1935 (2007).
104. Eltzschig, H. K. & Eckle, T. Ischemia and reperfusion–from mechanism to translation. *Nature Med.* **17,** 1391–1401 (2011).
105. Eltzschig, H. K. *et al.* Endothelial catabolism of extracellular adenosine during hypoxia: the role of surface adenosine deaminase and CD26. *Blood* **108,** 1602–1610 (2006).
106. Morote-Garcia, J. C., Rosenberger, P., Kuhlicke, J. & Eltzschig, H. K. HIF-1-dependent repression of adenosine kinase attenuates hypoxia-induced vascular leak. *Blood* **111,** 5571–5580 (2008).
107. Schingnitz, U. *et al.* Signaling through the A2B adenosine receptor dampens endotoxin-induced acute lung injury. *J. Immunol.* **184,** 5271–5279 (2010).
108. Eltzschig, H. K., Bonney, S. K. & Eckle, T. Attenuating myocardial ischemia by targeting A2B adenosine receptors. *Trends Mol. Med.* **19,** 345–354 (2013).
109. Eckle, T. *et al.* Adora2b-elicited Per2 stabilization promotes a HIF-dependent metabolic switch crucial for myocardial adaptation to ischemia. *Nature Med.* **18,** 774–782 (2012).
110. Eckle, T. *et al.* Crosstalk between the equilibrative nucleoside transporter ENT2 and alveolar Adora2b adenosine receptors dampens acute lung injury. *FASEB J.* **27,** 3078–3089 (2013).
111. Morote-Garcia, J. C., Rosenberger, P., Nivillac, N. M., Coe, I. R. & Eltzschig, H. K. Hypoxia-inducible factor-dependent repression of equilibrative nucleoside transporter 2 attenuates mucosal inflammation during intestinal hypoxia. *Gastroenterology* **136,** 607–618 (2009).
112. Frick, J. S. *et al.* Contribution of adenosine A2B receptors to inflammatory parameters of experimental colitis. *J. Immunol.* **182,** 4957–4964 (2009).
113. Csoka, B. *et al.* A2B adenosine receptors protect against sepsis-induced mortality by dampening excessive inflammation. *J. Immunol.* **185,** 542–550 (2010).
114. Belikoff, B. G. *et al.* A2B adenosine receptor blockade enhances macrophage-mediated bacterial phagocytosis and improves polymicrobial sepsis survival in mice. *J. Immunol.* **186,** 2444–2453 (2011).
115. Zhang, K. *et al.* Structure of the human $P2Y_{12}$ receptor in complex with an antithrombotic drug. *Nature* http://dx.doi.org/10.1038/nature13083 (23 March, 2014).
116. Zhang, J. *et al.* Agonist-bound structure of the human $P2Y_{12}$ receptor. *Nature* (in the press).

Supplementary Information is available in the online version of the paper.

Acknowledgements We acknowledge S. A. Eltzschig for help with artwork during manuscript preparation. The present research is supported by Deutsche Forschungsgemeinschaft grant ID7/3-1 ID7/4-1 and a grant by the Boehringer-Ingelheim Foundation to I.D., as well as National Institutes of Health grants R01-DK097075, R01-HL0921, R01-DK083385, R01-HL098294, POIHL114457-01 and a grant by the Crohn's and Colitis Foundation of America to H.K.E.

Author Contributions M.I., D.F. and H.K.E. all contributed to the writing of this paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.I. (marco.idzko@uniklinik-freiburg.de).

ARTICLE

doi:10.1038/nature13273

Total synthesis and isolation of citrinalin and cyclopiamine congeners

Eduardo V. Mercado-Marin[1]*, Pablo Garcia-Reynaga[1]*, Stelamar Romminger[2], Eli. F. Pimenta[2], David K. Romney[3], Michael W. Lodewyk[4], David E. Williams[5], Raymond J. Andersen[5], Scott J. Miller[3], Dean J. Tantillo[4], Roberto G. S. Berlinck[2] & Richmond Sarpong[1]

Many natural products that contain basic nitrogen atoms—for example alkaloids like morphine and quinine—have the potential to treat a broad range of human diseases. However, the presence of a nitrogen atom in a target molecule can complicate its chemical synthesis because of the basicity of nitrogen atoms and their susceptibility to oxidation. Obtaining such compounds by chemical synthesis can be further complicated by the presence of multiple nitrogen atoms, but it can be done by the selective introduction and removal of functional groups that mitigate basicity. Here we use such a strategy to complete the chemical syntheses of citrinalin B and cyclopiamine B. The chemical connections that have been realized as a result of these syntheses, in addition to the isolation of both 17-hydroxycitrinalin B and citrinalin C (which contains a bicyclo[2.2.2]diazaoctane structural unit) through carbon-13 feeding studies, support the existence of a common bicyclo[2.2.2]diazaoctane-containing biogenetic precursor to these compounds, as has been proposed previously.

The prenylated indole alkaloids are an emerging class of natural products typified by the presence of an indole ring, or derivatives thereof (that is, spirooxindole or pseudoindoxyl), decorated by one or more prenyl groups or the vestige of a prenyl group. Isolates from this family of natural products include citrinalins A and B (Fig. 1, **1** and **2**) and cyclopiamines A and B (**4** and **6**), which are the focus of this Article. The modifications of the indole core in the prenylated indole alkaloid family, which occur by a reaction with dimethylallyl pyrophosphate[1], results in the introduction of a chromene unit as is found in (+)stephacidin A (**10**; see blue highlighted portion) or a bicyclo[2.2.2]diazaoctane core that is typical of many congeners, including **11** and **12** (ref. 2) (see red highlighted portion).

Although structurally similar, the prenylated indole alkaloids display a diverse range of bioactivities including anti-tumour, insecticidal, anthelmintic, calmodulin-inhibition and antibacterial properties[3]. The recent discovery of citrinadins A (ref. 4) and B (ref. 5) (**7** and **8**) and PF1270A–PF1270C[6] (**9a**–**9c**) has added an unprecedented dimension to the structural motifs afforded by the *Penicillium* strains, and has raised several questions as to the biogenesis of these structurally related alkaloids. Recently, syntheses of citrinadins A and B have been achieved[7,8]. Particularly intriguing to us is a subset of this emerging subclass including citrinalins A and B (**1** and **2**) and cyclopiamines A and B (**4** and **6**), which, like the citrinadins, lack the bicyclo[2.2.2]diazaoctane framework and, remarkably, possess an alkyl nitro group. Cyclopiamines A and B (**4** and **6**) were discovered in 1979 in a toxinogenic strain of *Penicillium cyclopium*[9], whereas citrinalins A and B (**1** and **2**) were discovered in 2010 in a strain of *Penicillium citrinum*[10]. Although natural products that possess aryl nitro groups are known, those that contain aliphatic nitro groups are extremely rare[11]. As a result, the citrinalins and cyclopiamines, which also possess three nitrogen atoms in chemically distinct environments, are unusual and are therefore attractive targets for synthesis. The synthetic studies described here have culminated in the total syntheses of ent-citrinalin B (ent-**2**; ent, enantiomer) and cyclopiamine B (**6**), and, along with ^{13}C feeding studies that have resulted in the isolation of two new citrinalins, provide support for a proposed biogenesis of the subset of prenylated indole alkaloids that lack the bicyclo[2.2.2]diazaoctane core.

Biosynthetic connections

A stimulating connection may be drawn between cyclopiamine A and B via the intermediacy of nitronate iminium ion **5** (ref. 9) (Fig. 1). The interconversion of **4** and **6** was demonstrated by heating either compound in a mixture of dioxane and water or in dimethylformamide[9] (DMF). This led to a proposal that **6**, which is the more stable of the two isomers (we have computed **6** to be 9.6 kcal mol^{-1} lower in energy than **4** in a DMF solvent model; see Supplementary Information), may in fact be an isolation artefact. Given the likelihood that the citrinadins, citrinalins and cyclopiamines are all oxidative degradation products of a precursor containing a bicyclo[2.2.2]diazaoctane ring, such as marcfortine A (**11**; in the case of the citrinadins) or stephacidin A (**10**; in the case of the citrinalins and cyclopiamines), we wondered whether the citrinalins could be transformed to the cyclopiamines. On the basis of this assumption, it is particularly baffling that, unlike cyclopiamines A and B, which are related by an aza-Henry (or nitro-Mannich) reaction as shown in Fig. 1 (**4** ⇔ **6**, via **5**), citrinalin A and the originally proposed structure of citrinalin B (**3**) would be related not by the formal epimerization of the C22 stereocentre but rather by the nature of the relative configuration of the C14 carbon (highlighted in **2** and **3**). On the basis of the connection between cyclopiamines A and B demonstrated previously[9], we intuited that the structure of citrinalin B may be better represented by **2**. To support this proposal, we undertook a computational simulation of the ^{1}H and ^{13}C NMR spectra that would be expected for the neutral and salt forms of citrinalins A and B (Supplementary Information). As has been convincingly demonstrated in numerous cases, this method provides an accurate prediction of the structures of complex natural products[12]. We found that the computed and empirical data for the trifluoroacetic acid salt form of citrinalin A is in good agreement with those reported in ref. 10. The corrected mean absolute deviations

[1]Department of Chemistry, University of California, Berkeley, California 94720, USA. [2]Instituto de Quimica de Sao Carlos, Universidade de Sao Paulo, CP 780, CEP 13560-970, Sao Carlos, SP, Brazil. [3]Department of Chemistry, Yale University, PO Box 208107, New Haven, Connecticut 06520, USA. [4]Department of Chemistry, University of California, Davis, California 95616, USA. [5]Department of Chemistry and Earth and Ocean Sciences, University of British Columbia, Vancouver, British Columbia V6T IZl, Canada.
*These authors contributed equally to this work.

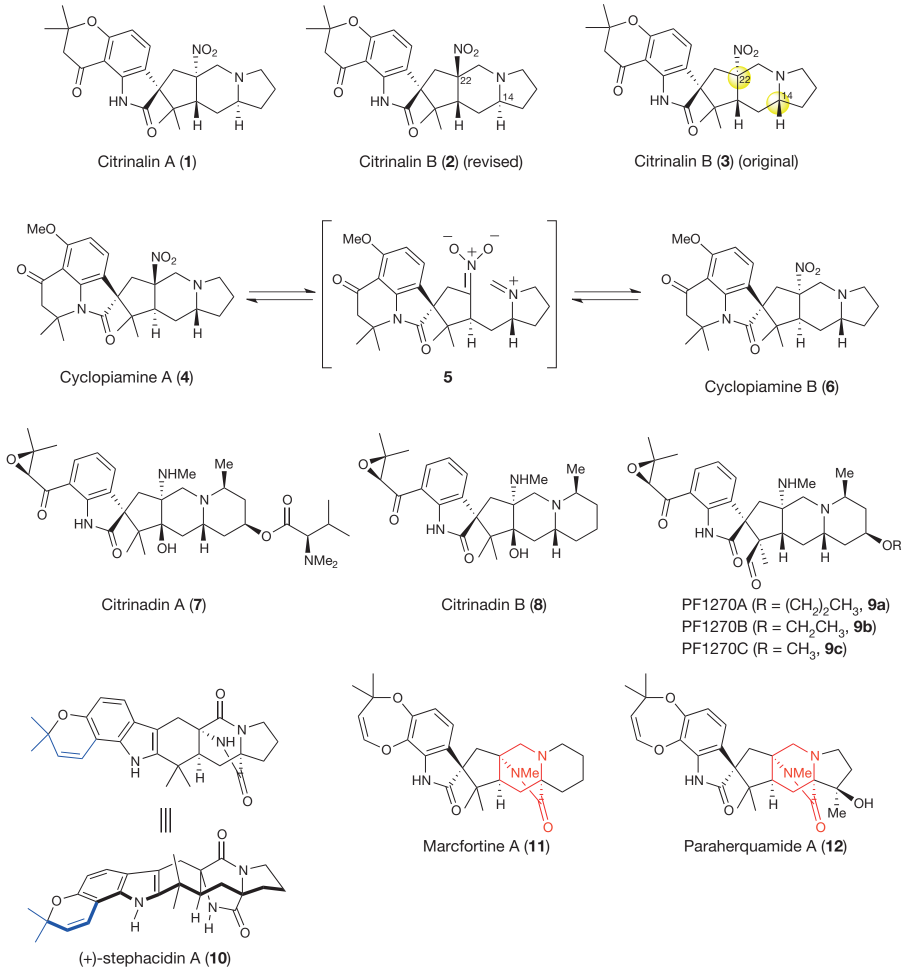


Figure 1 | Selected prenylated indole alkaloids. The prenylated indole alkaloid family encompasses over 80 natural products, some of which contain a bicyclo[2.2.2]diazaoctane core as in **10**, **11** and **12**. Recently, several members of this family (for example **1** and **4**) have emerged that do not possess this structural motif. Me, methyl.

(CMAD) in the ^{1}H and ^{13}C NMR resonances are 0.21 and 2.0 p.p.m., respectively (the largest outliers are 1.0 and 5.2 p.p.m., respectively). However, the computed data for the trifluoroacetic acid salt form of **3** (the originally proposed structure of citrinalin B) differs significantly from that recorded using the naturally occurring material (CMADs, 0.45 and 2.0 p.p.m.; largest outliers, 2.3 and 9.6 p.p.m. for ^{1}H and ^{13}C, respectively). The best match to the reported spectral data was found to correspond to **2** in its neutral form (CMADs, 0.12 and 1.6 p.p.m.; largest outliers, 0.38 and 4.4 p.p.m. for ^{1}H and ^{13}C, respectively), which corroborates the potentially similar biosynthetic connection that has been established for the cyclopiamines (outlined in Fig. 1). As a result, we chose to proceed on the assumption that **2** most probably represents the correct structure of citrinalin B. Ultimately, a reanalysis of the NMR data of citrinalin B, collected in MeOH-d_4 (Supplementary Information), corroborates the assignment of **2** as the true structure of citrinalin B.

Synthesis

As outlined in Fig. 2, cyclopiamine B (**6**) can be obtained from the enantiomer of citrinalin B (ent-**2**) by using a chromanone rearrangement to forge the tetrahydroquinolone structural moiety found in the cyclopiamines. In turn, ent-**2** could be taken back using an 'indole-to-spirooxindole' transform to fused hexacycle **13**. Fused indole **13** would

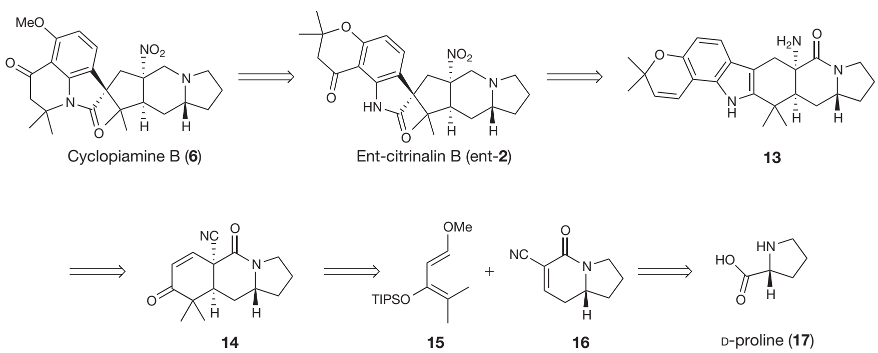


Figure 2 | Retrosynthetic analysis plan for cyclopiamine B and citrinalin B. The syntheses of natural products **2** and **6** are expected to arise from common intermediate **13**. TIPS, triisopropylsilyl.

arise from tricycle **14**, which may be prepared from diene **15**, the *t*-butyldimethylsilyl variant of which was first prepared in ref. 13, and tetrahydroindolizinone **16**, which would ultimately arise from D-proline (**17**).

We initiated our synthetic studies with the protection of D-proline by *t*-butoxycarbonyl (Fig. 3), which was followed by the reduction of the carboxylic acid group and Swern oxidation of the resulting hydroxyl to afford aldehyde **18** (ref. 14). Alkynylative homologation of the aldehyde group of **18** using the Ohira–Bestmann method[15], followed by removal of the *t*-butoxycarbonyl group and acylation with 2-cyanoacetyl chloride gives alkyne **19**. This serves as a substrate for an unprecedented formal cycloisomerization that probably proceeds via a metal vinylidene intermediate[16], anti-Markovnikov hydration and Knoevenagel condensation to give tetrahydroindolizinone **16**. At this stage, a $SnCl_4$-catalysed Diels–Alder [4+2] reaction[17] between **16** and diene **15,** and a subsequent basic work-up, affords an enone (not shown), which is iodinated to yield iodoenone **20** (ref. 18). A mild hydrolysis of the nitrile group of **20** is achieved using Pt-complex **21** (ref. 19) to afford the corresponding carboxamide, which serves as a substrate for a Hofmann rearrangement that is effected with phenyliodosyl bistrifluoroacetate to yield carbamate **22** (ref. 20). Suzuki cross-coupling of **22** with known boronic ester **23** (ref. 21) gives adduct **24**, which is efficiently converted to fused indole **25** using two sequential reductions—all in accord with the effective protocols established in ref. 22.

The face-selective oxygenation of C2/C3-fused indoles is a well-established route to hydroxyindolenines, which serve as precursors to the corresponding spirooxindoles[23]. We therefore reasoned that the oxygenation of indole **25** (Fig. 4a) could be a path to the spirooxindole structural moiety found in **1**, **2**, **4** and **6**. On the basis of related precedents for heteroatom-directed oxygenation[7,24,25], we expected the carbamate group of **25** to direct oxygenation to the α-face and provide **28**. Surprisingly, the use of Davis' oxaziridine[26] (**29**; 3.0 equiv.) leads to **26** and trace amounts of both hydroxyindolenine **28** and spirooxindole **27** (spirooxindole **27** arises via the intermediacy of hydroxyindolenine **26**). A survey of other oxaziridines, including **30** and **31**, leads, at best (using **31**), to a 1:1 ratio of the desired hydroxyindolenine, **28,** and both hydroxyindolenine **26** and spirooxindole **27**. Because the inherent face selectivity for the oxygenation of **25** is poor, attention was turned to the use of reagent control to achieve the desired diastereoselective oxygenation. In this regard, we were drawn to peptide-derived catalysts developed for oxygenations[27]. Following an investigation of a focused library of peptide catalysts, **32** (Fig. 4b) emerged as the superior catalyst (20 mol% loading) and provided hydroxyindolenine **28** in 83% yield from **25**. Hydroxyindolenine **28** rearranges with heating using $Sc(OTf)_3$ over 2 h to afford pseudoindoxyl **33** (Fig. 4c) instead of the desired spirooxindole. The equilibrium between pseudoindoxyls and spirooxindoles is well recognized and has been studied for the migration of C2 alkyl substituents[28] and C2 aryl substituents[29]. However, despite prolonged heating, further rearrangement of pseudoindoxyl **33** to the desired spirooxindole was not observed. It is possible that an intramolecular hydrogen bond stabilizes pseudoindoxyl **33** against further rearrangement (a bond distance of 2.24 Å is computed for the pseudoindoxyl carbonyl group and N–H proton of the carbamate group in **33**; see Supplementary Information). Evidence for a stabilizing intramolecular hydrogen bond in **33** comes from the observation that hydroxyindolenine **26** (prepared by oxidation of **25** with Davis' oxaziridine) rearranges readily at room temperature in the presence of mild acid to spirooxindole **27**; a pseudoindoxyl generated from **26** would lack the analogous stabilizing hydrogen bond. However, the possibility exists that **26** proceeds to an epoxide intermediate (see **A** in inset in Fig. 4c) that rearranges to **27**. The difficulty of further rearranging pseudoindoxyl **33** caused us to consider alternative approaches that would produce the desired spirooxindole structural moiety of the citrinalins and cyclopiamines.

Amino compound **35** (Fig. 5) was prepared on the assumption that an amino group, or some oxidized derivative thereof (for example the corresponding hydroxylamine), could serve as a hydrogen-bond donor to effect stereoselective oxygenation of the indole C2–C3 bond and then, by further oxidation to a nitroso or nitro group, remove the presumed intramolecular hydrogen bond that may stabilize the pseudoindoxyl form (as in **33**). It seemed reasonable that this sequence would facilitate the eventual conversion of **35** to nitro spirooxindole compound **36**. Initial experiments established that epoxidation of the chromene ring was a competing reaction that occurred under various oxygenation conditions. As such, we opted to effect a Wacker oxidation[30] of **25**

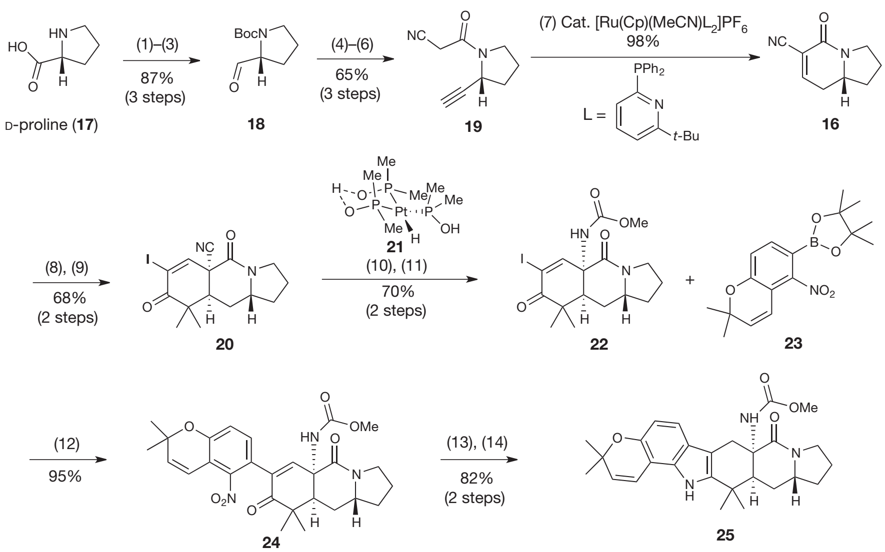


Figure 3 | Preparation of fused hexacycle 25. The use of a Diels–Alder reaction involving a proline-derived indolizidinone dienophile affords a key tricycle that is advanced to hexacycle **25** by Suzuki coupling to boronic ester **23**. Reagents and conditions are as follows: (1) di-*t*-butyl dicarbonate (Boc_2O), $NaHCO_3$, H_2O and tetrahydrofuran (THF), room temperature (RT = 23 °C); (2) BH_3•THF, THF, 0 °C to RT; (3) $(COCl)_2$, dimethylsulphoxide, CH_2Cl_2, diisopropylethylamine, −78 °C; (4) dimethyl (diazomethyl)phosphonate, K_2CO_3, MeOH, 0 °C to RT; (5) 4N HCl and dioxane, 0 °C to RT; (6) 2-cyanoacetylchloride, Et_3N, CH_2Cl_2, 0 °C to RT; (7) acetonitrile bis[2-diphenylphosphino-6-*t*-butylpyridine] cyclopentadienylruthenium(II) hexafluorophosphate (8 mol%), acetone and H_2O, 70 °C; (8) **15**, $SnCl_4$, −78 to −42 °C; (9) I_2, 4-dimethylaminopyridine, pyridine and CCl_4, 60 °C; (10) **21** (20 mol%), EtOH and H_2O, RT; (11) phenyliodosyl bistrifluoroacetate, MeOH, RT; (12) dppfPdCl$_2$ (10 mol%), K_3PO_4, DMF, 40 °C; (13) Zn dust, NH_4Cl, HCO_2NH_4, *p*-TsOH, MeOH, RT; (14) $NaCNBH_3$, 1N HCl(aq.), 0 °C to RT. dppf, (diphenylphosphino)ferrocene; Et, ethyl; *t*-Bu, *t*-Butyl; Ts, tosyl.

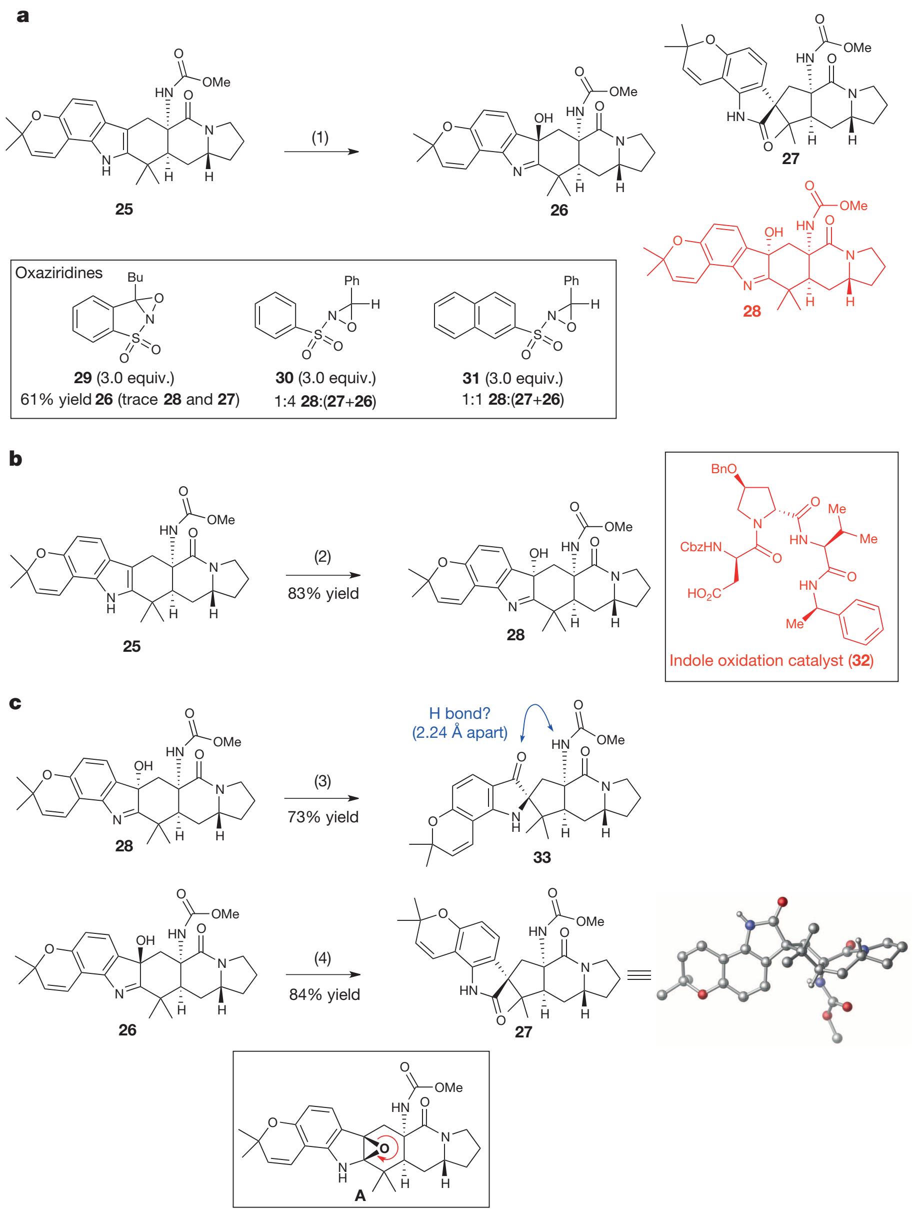


Figure 4 | Face-selective oxygenation of fused hexacycle 25. a, Oxidative rearrangement studies of fused indole **25** with a range of oxaziridines leads predominantly to the undesired, epimeric, hydroxyindolenine (**26**) and spirooxindole (**27**). **b**, Use of indole oxidation peptide catalyst **32** to effect oxidation yields the desired hydroxyindolenine (**28**). **c**, Hydroxyindolenine **28** rearranges to an undesired pseudoindoxyl (**33**), whereas the epimeric hydroxyindolenine (**26**) affords the corresponding spirooxindole (**27**). Reagents and conditions are as follows: (1) oxaziridine (**29**, **30**, or **31**), CH_2Cl_2, RT; (2) **32** (20 mol%), 4-dimethylaminopyridine, diisopropylcarbodiimide, H_2O_2, $CHCl_3$, 4 °C; (3) $Sc(OTf)_3$, toluene, 110 °C; (4) 23 mM HCl, CH_2Cl_2, RT. Bn, benzyl; Cbz, carboxybenzyl; Ph, phenyl.

to afford chromanone **34** (Fig. 5), which would be advantageous because the chromanone unit is found in the citrinalins and cyclopiamines. Remarkably, treatment of **35** (following removal of the methoxycarbonyl group in **34**) with an excess of dimethyldioxirane (formed *in situ* from acetone and Oxone) affords spirooxindole **36** as the major product (diastereomeric ratio, 4:1) where the spiro centre is as desired and the nitro group has been installed. Studies of dimethyldioxirane oxidations of indoles to spirooxindoles[31] suggest that spirooxindole **36** might arise from epoxide **B** (Fig. 5, inset). Therefore, it is possible that the introduction of the chromanone diminishes the participatory role of the indole nitrogen lone pair leading, after rearrangement (see direction of arrow in **B**), to **36**. With spirooxindole **36** in hand, what remained was a selective removal of the tertiary amide carbonyl group by reduction, which had to be accomplished in the presence of the chromanone and secondary amide carbonyl groups as well as the newly introduced nitro group. After extensive investigation, this task was effectively accomplished using a modification of a known procedure[32] by treating **36** with a variant of Meerwein's salt (Me_3OBF_4), which probably leads to a methylated amidinium intermediate that is cleanly reduced with sodium cyanoborohydride to give ent-citrinalin B (ent-**2**) in 66% yield (79% based on recovered starting material). The spectroscopic data for the neutral form of ent-**2** are fully consistent with previous data reported for the compound believed to be citrinalin B (ref. 10; corroborating the computational predictions and reanalysis in MeOH-d_4), except for the sign of optical rotation, which is opposite. The structure of ent-**2** was unambiguously confirmed by X-ray crystallographic analysis of its HCl salt. Ent-citrinalin B is easily converted to cyclopiamine B (**6**) on treatment of ent-**2** with sodium hydride and heating (to effect the conversion of chromanone to tetrahydroquinolone) and subsequent methylation of the resulting phenol. The structure of cyclopiamine B (**6**) was also unambiguously confirmed by X-ray

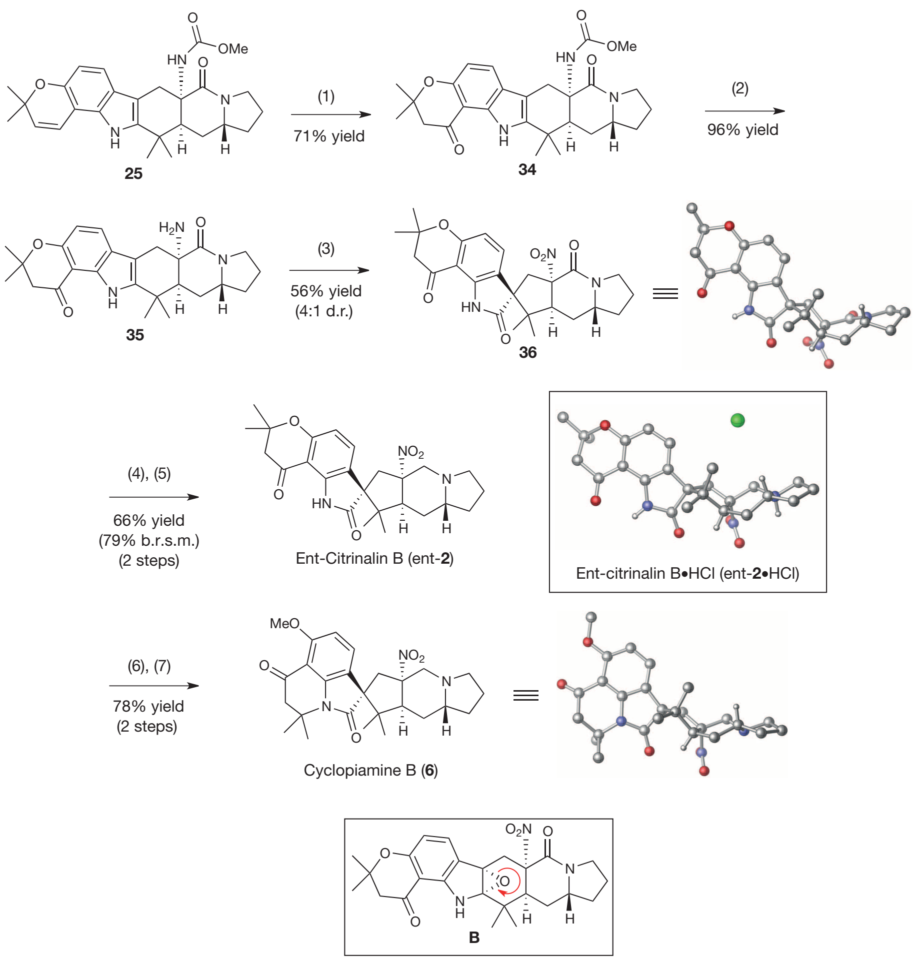


Figure 5 | Completion of the syntheses of ent-citrinalin B and cyclopiamine B. The total syntheses of **2** and **6** required the identification of conditions that accomplished the oxidation of the amino group and spirooxindole formation in one pot as well as unique conditions for the selective reduction of the tertiary amide carbonyl group. The rearrangement of ent-citrinalin B (**2**) to cyclopiamine B (**6**) was also demonstrated. Reagents and conditions are as follows: (1) $Pd(OAc)_2$ (40 mol%), benzoquinone, H_2SO_4, MeCN and H_2O, RT; (2) Me_2S, methanesulphonic acid, 40 °C; (3) Oxone (10 equiv.), $NaHCO_3$, acetone and H_2O, 0 °C to RT; (4) Me_3OBF_4, CH_2Cl_2, 4 Å molecular sieves, 45 °C; (5) $NaCNBH_3$, MeOH, 0 °C; (6) NaH, DMF, 60 °C; (7) MeI, K_2CO_3, acetone, 60 °C. b.r.s.m., based on recovered starting material; d.r., diastereomeric ratio; Oxone, potassium peroxymonosulphate.

crystallographic analysis. Thus, the synthesis of ent-**2** and its conversion to **6** show that ent-**2** is the true structure of citrinalin B, albeit the enantiomer of the naturally occurring material.

Biosynthetic considerations

The total syntheses of ent-citrinalin B (ent-**2**; 19 steps from D-proline, 5.5% overall yield) and cyclopiamine B (**6**; 21 steps from D-proline, 4.3% overall yield) not only unambiguously establish the structures of these metabolites, but also provide possible insight into the biogenesis of these natural products (especially as to the possible formation of the cyclopiamines from the citrinalins).

The citrinalins, and in turn the cyclopiamines, probably arise from a bicyclo[2.2.2]diazaoctane precursor. However, such a precursor was unknown before the findings that are reported herein (see below). Consistent with numerous biosynthetic studies of the prenylated indole alkaloids, the structural features of **1**, **2**, **4** and **6** suggest that tryptophan, proline and two isoprene units are biosynthetic precursors to these compounds. Although no biosynthetic studies on **1** and **2** or **4** and **6** or the related citrinadins and PF1270 alkaloids has appeared, it has been suggested that they are derived from bicyclo[2.2.2]diazaoctane precursors that suffer the 'loss' of one diketopiperazine carbonyl group[5]. Through the isolation of 17-hydroxycitrinalin B (**37**; Fig. 6a) and, more importantly, citrinalin C (**38**) following a series of stable isotope labelling experiments (summarized in Fig. 6b; see Supplementary Information), we have now obtained support for the possible biogenesis of the citrinalins and cyclopiamines from a precursor bearing the bicyclo[2.2.2]diazaoctane moiety.

The NMR and mass spectroscopy characterization data for **37** is fully consistent with the assigned structure. Moreover, the assigned relative configuration fully corroborates the revised structure of citrinalin B (**2**).

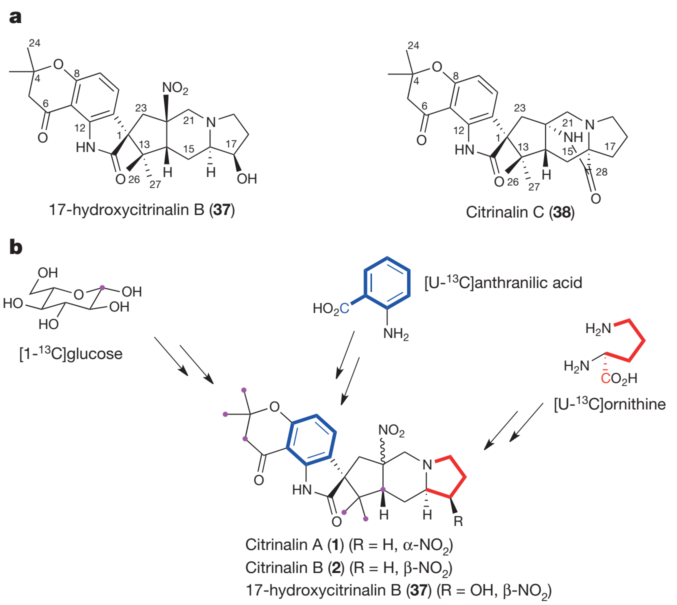


Figure 6 | Isolation of two new citrinalins and ^{13}C labelling studies. **a**, Structures of 17-hydroxycitrinalin B and citrinalin C. Two additional citrinalins, **37** and **38**, were isolated on refractionation and reanalysis of secondary metabolites from *P. citrinum* F53. **b**, Summary of the ^{13}C labelling studies. ^{13}C incorporation studies of *P. citrinum* F53 reveal that glucose (pink), anthranilic acid (blue) and ornithine (red) are biosynthetic precursors to the citrinalins.

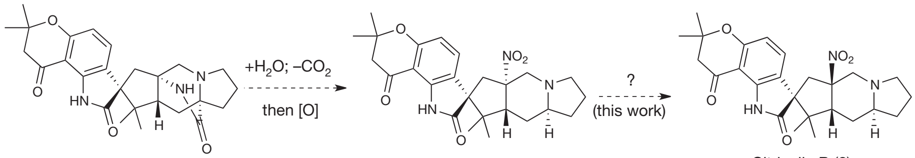


Figure 7 | Biosynthetic proposal for citrinalins. Consistent with previous reports on the bicyclo[2.2.2]diazaoctane congeners, the citrinalins probably arise through an intramolecular Diels–Alder reaction to form citrinalin C (**38**), which is followed by a decarboxylation event and amine-group oxidation to the nitro group.

By analogy to citrinalin B (**2**), the absolute configuration of **37** was assigned as 1*S*,14*R*,16*R*,17*R*,22*R*. 17-hydroxycitrinalin B (**37**) was initially isolated from *P. citrinum* F53 grown in a nitrogen-depleted culture medium. Stable isotope feeding studies with [U-^{13}C]anthranilic acid and [1-^{13}C]glucose gave significant ^{13}C labelling (Supplementary Information). High levels of [U-^{13}C]ornithine were also incorporated into **37,** and additional feeding studies with [U-^{13}C]proline gave almost undetectable labelling. Ornithine is a well-known biosynthetic precursor to proline, but to our knowledge it has never been reported as an efficient substrate for isotopic labelling of the putative proline-derived atoms in the biosynthesis of prenylated indole alkaloids of fungal origin bearing the bicyclo[2.2.2]diazaoctane moiety. The labelling investigations suggest that 17-hydroxycitrinalin B (**37**) might arise from either 3-hydroxyl ornithine, 3-hydroxy proline or by the late-stage oxygenation of the citrinalin A, B or C skeleton.

Citrinalin C (**38**), isolated as a minor component from the culture medium of *P. citrinum* F53, gives NMR and mass spectroscopic data (Supplementary Table 4) that is fully consistent with the relative and absolute configuration illustrated for this natural product. The isolation of **38**, along with the congeners lacking the bicyclo[2.2.2]diazaoctane structural moiety from *P. citrinum* F53, lends support to a bicyclo[2.2.2] diazaoctane-containing precursor, which arises from a committed intramolecular Diels–Alder cycloaddition step such as that studied in detail for other congeners[33]. Hydrolysis of the amide bridge of citrinalin C (**38**; Fig. 7), followed by decarboxylation, and amino-group oxidation to the nitro group, as proposed in the biosynthesis of the structurally related citrinadin B[5], would then yield citrinalin A. These latter steps are the subject of current biosynthesis studies.

A question that remained at this stage concerned the biogenesis of citrinalin B. On the basis of observations of the cyclopiamine series[9] (see **4** → **6** in Fig. 1), we anticipated that citrinalin A (**1**) might be converted to citrinalin B (**2**) via a nitronate iminium intermediate analogous to **5**. In the event, heating a solution of a naturally occurring sample of citrinalin A (**1**) in DMF-d_7 at 100 °C for 20 h leads to a 1:1 ratio of **1** and **2** (with complete conversion to citrinalin B (**2**) after 60 h; see Supplementary Fig. 22), confirming the connection of these metabolites presumably by the same aza-Henry or nitro-Mannich epimerization sequence established for the cyclopiamines[9]. However, we have observed some key differences. First, the epimerization in the citrinalin series occurs at a qualitatively lower rate (probably owing to a non-productive proton transfer from the vinylogous imide N–H to the tertiary amine) and higher temperature. In addition, we have not been able to achieve any observable conversion of citrinalin B to citrinalin A even at elevated temperatures (165 °C) over prolonged periods (24 h). Our current efforts are focused on gaining a deeper understanding of these differences and exploring the biosynthetic conversion of citrinalin C to citrinalin A.

Conclusion

We have reported the total syntheses of the prenylated indole alkaloids ent-citrinalin B and cyclopiamine B. Our results unambiguously identify citrinalin B through synthesis, a reanalysis of the naturally isolated material and an X-ray crystallographic study. Our studies on the isolation of metabolites from *P. citrinum* suggest that a bicyclo[2.2.2]diazaoctane-containing metabolite such as citrinalin C (**38**) is an intermediate in the biogenesis of citrinalins A (**1**) and B (**2**) (Fig. 7). The extension of the synthetic methods reported here to the syntheses of other prenylated indole alkaloids is ongoing and will be reported in due course.

METHODS SUMMARY

All reactions were performed under a nitrogen atmosphere using dry solvents under anhydrous conditions, unless otherwise noted. Dry tetrahydrofuran, toluene, methanol, triethylamine, benzene and diethyl ether were obtained by passing the commercially available, oxygen-free solvents through activated alumina columns from GlassContour. Dichloromethane was distilled over calcium hydride under a nitrogen atmosphere. Yields refer to materials purified using silica gel column chromatography. Full experimental details and characterization data for all new compounds (^{1}H NMR, ^{13}C NMR, mass spectrometry, infrared, R_f value), including **14–36**, **2** and **6**, appear in Supplementary Information. Crystallographic data were collected on a MicroSTAR-H APEX II (ChexStar: RUA #1091) instrument, and the Bruker SAINT and SADABS software programs were used for integrating and scaling the data, respectively. The CYLVIEW program (developed by C. Y. Legault) was used for X-ray depictions. Computational analyses were conducted following conformational searches using the MMFF94 force field (SPARTAN'10). Density functional theory calculations were performed with GAUSSIAN09 (B3LYP/6-31+G(d,p) theory level). Full details are included in Supplementary Information.

Received 29 January; accepted 21 March 2014.

1. Stocking, E. M., Sanz-Cervera, J. F. & Williams, R. M. Reverse versus normal prenyl transferases in paraherquamide biosynthesis exhibit distinct facial selectivities. *Angew. Chem. Int. Ed.* **38,** 786–789 (1999).
2. Finefield, J. M., Frisvad, J. C., Sherman, D. H. & Williams, R. M. Fungal origins of the bicyclo[2.2.2]diazaoctane ring system of prenylated indole alkaloids. *J. Nat. Prod.* **75,** 812–833 (2012).
3. Miller, K. A. & Williams, R. M. Synthetic approaches to the bicyclo[2.2.2]diazaoctane ring system common to the paraherquamides, stephacidins and related prenylated indole alkaloids. *Chem. Soc. Rev.* **38,** 3160–3174 (2009).
4. Tsuda, M. *et al.* Citrinadin A, a novel pentacyclic alkaloid from marine-derived fungus *Penicillium citrinum*. *Org. Lett.* **6,** 3087–3089 (2004).
5. Mugishima, T. *et al.* Absolute stereochemistry of citrinadins A and B from marine-derived fungus. *J. Org. Chem.* **70,** 9430–9435 (2005).
6. Kushida, N. *et al.* PF1270A, B and C, novel histamine H3 receptor ligands produced by *Penicillium waksmanii* PF1270. *J. Antibiot. (Tokyo)* **60,** 667–673 (2007).
7. Bian, Z., Marvin, C. C. & Martin, S. F. Enantioselective total synthesis of (−)-citrinadin A and revision of its stereochemical structure. *J. Am. Chem. Soc.* **135,** 10886–10889 (2013).
8. Kong, K. *et al.* An enantioselective total synthesis and stereochemical revision of (+)-citrinadin B. *J. Am. Chem. Soc.* **135,** 10890–10893 (2013).
9. Bond, R. F., Boeyens, J. C. A., Holzapfel, C. W. & Steyn, P. S. Cyclopiamines A and B, novel oxindole metabolites of *Penicillium cyclopium* westling. *J. Chem. Soc. Perkin Trans. I* 1751–1761 (1979).
10. Pimenta, E. F. *et al.* Use of experimental design for the optimization of the production of new secondary metabolites by two penicillium species. *J. Nat. Prod.* **73,** 1821–1832 (2010).
11. Parry, R., Nishino, S. & Spain, J. Naturally-occurring nitro compounds. *Nat. Prod. Rep.* **28,** 152–167 (2011).
12. Lodewyk, M. W., Siebert, M. R. & Tantillo, D. J. Computational prediction of ^{1}H and ^{13}C chemical shifts: a useful tool for natural product, mechanistic and synthetic organic chemistry. *Chem. Rev.* **112,** 1839–1862 (2012).
13. Jewett, J. C. & Rawal, V. H. Total synthesis of pederin. *Angew. Chem. Int. Ed.* **46,** 6502–6504 (2007).
14. Omura, K. & Swern, D. Oxidation of alcohols by "activated" dimethyl sulfoxide. A preparative, steric and mechanistic study. *Tetrahedron* **34,** 1651–1660 (1978).
15. Ohira, S. Methanolysis of dimethyl (1-diazo-2-oxopropyl) phosphonate: generation of dimethyl (diazomethyl) phosphonate and reaction with carbonyl compounds. *Synth. Commun.* **19,** 561–564 (1989).
16. Grotjahn, D. B. & Lev, D. A. A general bifunctional catalyst for the anti-Markovnikov hydration of terminal alkynes to aldehydes gives enzyme-like rate and selectivity enhancements. *J. Am. Chem. Soc.* **126,** 12232–12233 (2004).

17. Kishi, Y. *et al.* Synthetic approach towards tetrodotoxin. I. Diels-Alder reaction of alpha-oximinoethylbenzoquinones with butadiene. *Tetrahedr. Lett.* **11,** 5127–5128 (1970).
18. Johnson, C. R. *et al.* Direct alpha-iodination of cycloalkenones. *Tetrahedr. Lett.* **33,** 917–918 (1992).
19. Ghaffar, T. & Parkins, A. W. A new homogeneous platinum containing catalyst for the hydrolysis of nitriles. *Tetrahedr. Lett.* **36,** 8657–8660 (1995).
20. Moriarty, R. M., Chany, C. J. II, Vaid, R. K., Prakash, O. & Tuladhar, S. M. Preparation of methyl carbamates from primary alkyl- and arylcarboxamides using hypervalent iodine. *J. Org. Chem.* **58,** 2478–2482 (1993).
21. Herzon, S. B. & Myers, A. G. Enantioselective synthesis of stephacidin B. *J. Am. Chem. Soc.* **127,** 5342–5344 (2005).
22. Myers, A. G. & Herzon, S. B. Identification of a novel Michael acceptor group for the reversible addition of oxygen- and sulfur-based nucleophiles. Synthesis and reactivity of the 3-alkylidene-3*H*-indole 1-oxide function of avrainvillamide. *J. Am. Chem. Soc.* **125,** 12080–12081 (2003).
23. Marti, C. & Carreira, E. Construction of spiro[pyrrolidine-3,3′-oxindoles] − recent applications to the synthesis of oxindole alkaloids. *Eur. J. Org. Chem.* 2209–2219 (2003).
24. Guerrero, C. A. & Sorensen, E. J. Concise, stereocontrolled synthesis of citrinadin B core architecture. *Org. Lett.* **13,** 5164–5167 (2011).
25. Grubbs, A. W., Artman, G. D. III, Tsukamoto, S. & Williams, R. M. A concise total synthesis of the notoamides C and D. *Angew. Chem. Int. Ed.* **46,** 2257–2261 (2007).
26. Davis, F. A. & Stringer, O. D. Chemistry of oxaziridines. 2. Improved synthesis of 2-sulfonyloxaziridines. *J. Org. Chem.* **47,** 1774–1775 (1982).
27. Kolundzic, F. *et al.* Chemoselective and enantioselective oxidation of indoles employing aspartyl peptide catalysts. *J. Am. Chem. Soc.* **133,** 9104–9111 (2011).
28. Güller, R. & Borschberg, H.-J. A stereoselective transformation of pseudoindoxyls into oxindoles in a single operation. *Tetrahedr. Lett.* **35,** 865–868 (1994).
29. Movassaghi, M., Schmidt, M. A. & Ashenhurst, J. A. Stereoselective oxidative rearrangements of 2-aryl tryptamine derivatives. *Org. Lett.* **10,** 4009–4012 (2008).
30. Miller, D. G. & Wayner, D. D. M. Improved method for the Wacker oxidation of cyclic and internal olefins. *J. Org. Chem.* **55,** 2924–2927 (1990).
31. Zhang, X. & Foote, C. S. Dimethyldioxirane oxidation of indole derivatives. Formation of novel indole-2,3-epoxides and a versatile synthetic route to indolinones and indolines. *J. Am. Chem. Soc.* **115,** 8867–8868 (1993).
32. Borch, R. F. A new method for the reduction of secondary and tertiary amides. *Tetrahedr. Lett.* **9,** 61–65 (1968).
33. Ding, Y. *et al.* Genome-based characterization of two prenylation steps in the assembly of the stephacidin and notoamide anticancer agents in a marine-derived *Aspergillus* sp. *J. Am. Chem. Soc.* **132,** 12733–12740 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements R.S. and P.G.-R. thank the US National Institutes of Health (NIH; NIGMS RO1 086374) for financial support. R.S. is a Camille Dreyfus Teacher Scholar. E.V.M.-M. acknowledges the US National Science Foundation (NSF) for a graduate fellowship (GRFP). S.R., E.F.P. and R.G.S.B. are grateful to the Brazilian National Council of Technological and Scientific Development (CNPq; grant 470643/2010-2) and the São Paulo Research Foundation (FAPESP; grant 2012/50026-3) for funding. D.E.W. and R.J.A. thank NSEPC for funding. M.W.L. and D.J.T. acknowledge support from the NSF (CHE-0957416 and supercomputing resources through a grant from the XSEDE programme: CHE-030089). S.J.M. is grateful to the NIH for support (GM096403). We thank A. DiPasquale for solving the crystal structures of ent-**2**•HCl, **6**, **27** and **36** (supported by NIH Shared Instrumentation Grant S10-RR027172). We would like to thank T. Lebold, R. M. Williams and D. Sherman for discussions.

Author Contributions R.S. conceived and directed the synthetic aspects of the research, and wrote the majority of the manuscript (with input from all authors) except the section on biosynthesis, which was contributed by S.R., E.F.P., D.E.W., R.J.A. and R.G.S.B. The synthetic plan was designed by R.S. with input from E.V.M.-M. and P.G.-R. who carried out the plan under the supervision of R.S. Oxidation catalyst **32** was provided by D.K.R. and S.J.M., who, along with P.G.-R., E.V.M.-M. and R.S., designed the oxidation studies of **25**, which were done by P.G.-R. The computational NMR studies of **1**, **2** and **3** were designed and performed by M.W.L. and D.J.T. with input from P.G.-R., E.V.M.-M. and R.S. Biosynthetic studies were designed and conducted by S.R., E.F.P. and R.G.S.B., who also isolated and characterized **3**, **37** and **38**. D.E.W. and R.J.A. provided facilities and contributed to the purification, data analysis and structural analysis of **3**, **37** and **38**.

Author Information A sample of the *P. citrinum* strain F53 is deposited at the Brazilian Collection of Environmental and Industrial Microorganisms under the accession code CBMAI 1186. Crystallographic data for crystal structures ent-**2**•HCl, **6**, **27** and **36** have been deposited at the Cambridge Crystallographic Data Centre (http://www.ccdc.cam.ac.uk) under accession codes CCDC 984477, CCDC 984478, CCDC 984480 and CCDC 984479, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.G.S.B. (rgsberlinck@iqsc.usp.br) or R.S. (rsarpong@berkeley.edu).

doi:10.1038/nature13307

Galanin neurons in the medial preoptic area govern parental behaviour

Zheng Wu[1], Anita E. Autry[1], Joseph F. Bergan[1], Mitsuko Watabe-Uchida[2] & Catherine G. Dulac[1]

Mice display robust, stereotyped behaviours towards pups: virgin males typically attack pups, whereas virgin females and sexually experienced males and females display parental care. Here we show that virgin males genetically impaired in vomeronasal sensing do not attack pups and are parental. Furthermore, we uncover a subset of galanin-expressing neurons in the medial preoptic area (MPOA) that are specifically activated during male and female parenting, and a different subpopulation that is activated during mating. Genetic ablation of MPOA galanin neurons results in marked impairment of parental responses in males and females and affects male mating. Optogenetic activation of these neurons in virgin males suppresses inter-male and pup-directed aggression and induces pup grooming. Thus, MPOA galanin neurons emerge as an essential regulatory node of male and female parenting behaviour and other social responses. These results provide an entry point to a circuit-level dissection of parental behaviour and its modulation by social experience.

Understanding how neural circuits drive social behaviour is a fundamental question in neuroscience. Parental interactions aimed at the care and protection of young are essential for the survival of offspring in many animal species. Elaborate parental behaviour is a defining feature of mammals, presumably regulated by evolutionarily conserved neural circuits[1]. Intriguingly, the respective roles of the two parents in offspring care differ across highly related species: whereas mothers usually assume the largest share of parenting, the contribution of fathers varies markedly between species, ranging from dedicated parenting of pups to neglect and aggression[2,3]. The identification of neuronal circuits controlling the display of parental behaviour in males and females should help elucidate neural mechanisms underlying this essential social behaviour and provide novel insights into the regulation of sexually dimorphic brain functions.

Insights into the neurobiology of parental behaviour come primarily from studies in rodents[1]. Virgin rats find foreign pups aversive but exhibit parental care after continuous exposure to the pups[4], or after priming with hormones characteristic of parturient females[5,6]. In laboratory mice, virgin males and females exhibit markedly different behaviours towards pups. Virgin males typically attack pups[7,8], whereas virgin females exhibit spontaneous, stereotyped displays of maternal care[2,7]. Remarkably, males stop attacking pups and transiently become paternal after mating, starting near the time of birth of the pups and lasting until weaning[9–11]. In female rats, the medial preoptic area (MPOA) and the dopaminergic system have been implicated in the control of maternal behaviour[12,13]. However, the neural mechanisms underlying distinct parental behaviours in females and males with different social experience remain unknown.

Vomeronasal control of pup-directed aggression

The vomeronasal system plays an essential role in regulating sex-specific behaviours[14]. Males with impaired vomeronasal organ (VNO) signalling mount males and females, suggesting impaired gender identification[15]. Further, VNO-deficient females show notable male-like mounting and courtship displays, suggesting that the vomeronasal pathway constitutively represses male-specific behaviour circuits in females[16]. We proposed that, in males, the vomeronasal pathway may similarly regulate female-typical behaviours such as parenting. This idea is supported by evidence that vomeronasal areas are activated during pup-directed aggression and that disrupted VNO signalling in males reduces aggression and facilitates parenting[17–19].

We used genetic tools to confirm the role of VNO inputs in pup-directed behaviours. Genetic ablation of TRPC2, a VNO-specific ion channel, impairs vomeronasal signalling[15,20]. Adult *Trpc2*$^{-/-}$ virgin males and females and *Trpc2*$^{+/-}$ littermates were presented with C57BL/6J pups and behavioural responses were observed. In contrast to *Trpc2*$^{+/-}$ littermates, *Trpc2*$^{-/-}$ virgin males showed marked reductions in pup-directed aggression (Fig. 1a). Furthermore, a large fraction of *Trpc2*$^{-/-}$ virgin males exhibited parental care typical of females and fathers (Fig. 1a). Quantification of behaviour towards pups showed that *Trpc2*$^{-/-}$ males retrieved pups with shorter latency, engaged in more nest-building, and were in the nest crouching over and grooming pups longer than *Trpc2*$^{+/-}$ males. *Trpc2*$^{-/-}$ males, although clearly parental, displayed less parenting than *Trpc2*$^{-/-}$ females (Fig. 1b–f).

We next investigated the post-mating switch from attacking pups to paternal behaviour originally reported in the CF1 mouse strain[11]. Virgin control and mated males tested 1–2 days, or 10–12 days after mating attacked pups. However, mated males tested just before pups were born at day 17–20 did not attack pups, with half displaying paternal behaviour. All males tested at day 25–27 were paternal, consistent with previous studies[11,18,21] (Fig. 1g).

Thus, opposing behaviour circuits co-exist in the male brain to regulate pup-directed aggression and parenting behaviours according to social context. In virgin males, vomeronasal circuits activated by pup cues elicit pup-directed aggression while pathways underlying parenting behaviour remain silent. By contrast, mated males repress VNO-evoked aggression and instead activate parenting circuits.

Neuronal activation during parenting

To identify the brain regions involved in parental care, we compared the brain activity patterns of virgin males versus virgin females and paternal males using induction of the immediate early gene *c-fos* (also known as *Fos*) as a read-out of neuronal activation after exposure to pups. We focused our analysis on the hypothalamus, amygdala and other regions involved in social behaviours (Methods).

[1]Howard Hughes Medical Institute, Department of Molecular and Cellular Biology, Center for Brain Science, Harvard University, Cambridge, Massachusetts 02138, USA. [2]Department of Molecular and Cellular Biology, Center for Brain Science, Harvard University, Cambridge, Massachusetts 02138, USA.

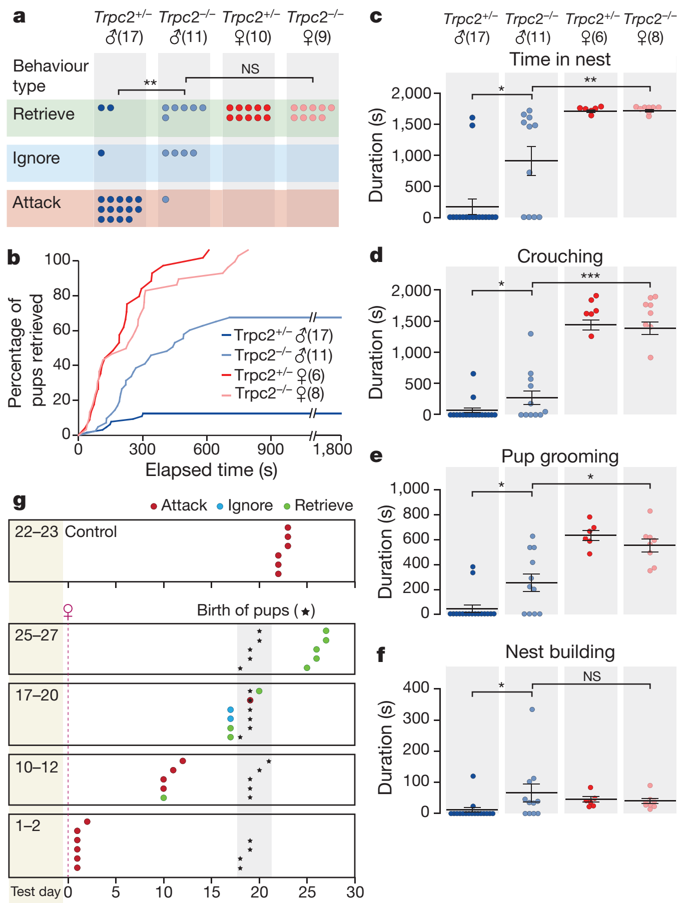


Figure 1 | Pup-directed behaviour of *Trpc2*$^{-/-}$ and *Trpc2*$^{+/-}$ virgin animals and switch from attack to parenting in males after mating. **a**, Behaviour analysis of *Trpc2*$^{-/-}$ and *Trpc2*$^{+/-}$ virgin males demonstrates significantly different responses to pups in the presence or absence of VNO signalling. Chi-square test with Bonferroni correction, $**P < 0.01$. **b**, Combined percentage of pups (out of four) retrieved by an animal group as a function of time. Kolmogorov–Smirnov test with Bonferroni correction, $P < 0.001$ between *Trpc2*$^{-/-}$ and *Trpc2*$^{+/-}$ males, $P < 0.01$ between *Trpc2*$^{-/-}$ males and *Trpc2*$^{-/-}$ females. **c–f**, Time spent in nest (**c**), and duration of crouching (**d**), pup grooming (**e**) and nest building (**f**). Mean ± s.e.m.; Mann–Whitney test with Bonferroni correction, $*P < 0.05$, $**P < 0.01$, $***P < 0.001$; NS, not significant. **g**, Behaviour of *Trpc2*$^{+/-}$ males tested after increasing durations of cohabitation with females subsequent to mating. Males mated on day 0 except virgin controls, which were individually housed from day 0 throughout the test. Male behaviour switches from attack to parenting at a time period after mating that corresponds to the birth of their pups.

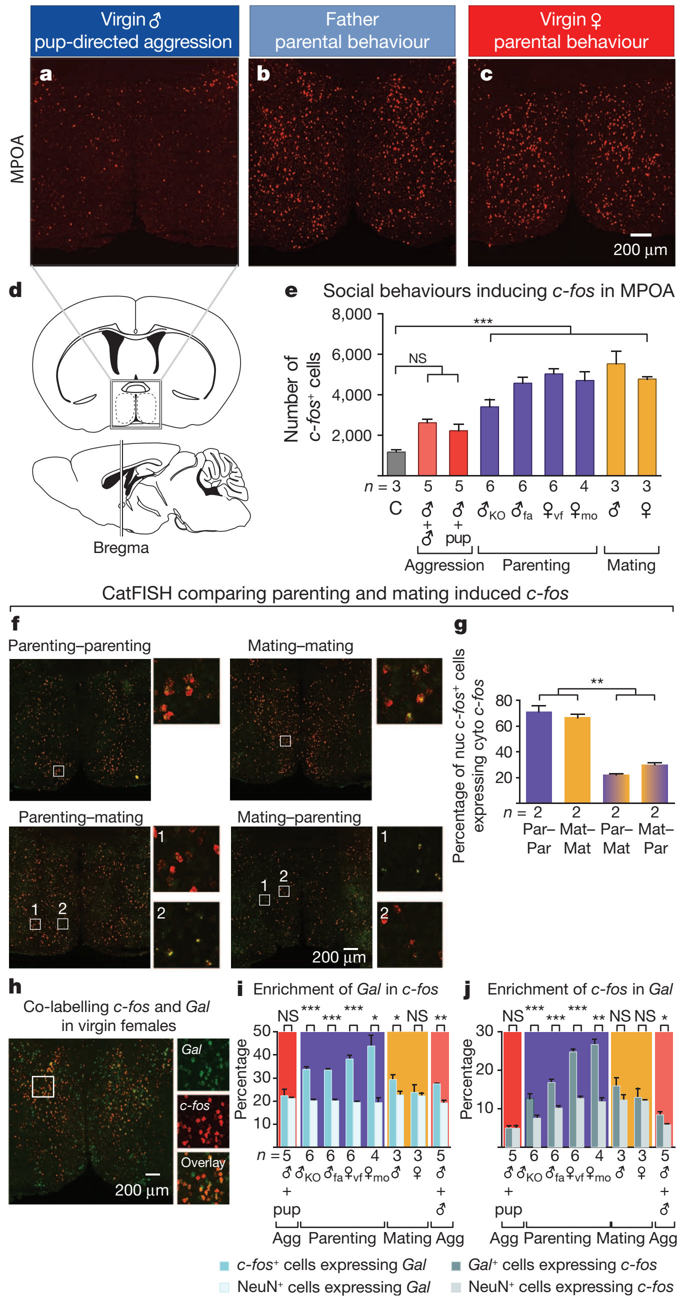


Figure 2 | Parenting activates galanin-expressing neurons in the MPOA. **a–c**, *c-fos* mRNA expression in the MPOA of virgin males (**a**), fathers (**b**) and virgin females (**c**) after interaction with pups. **d**, Schematic illustration of the MPOA in sagittal and coronal sections, adapted from the Paxinos and Franklin mouse brain atlas. **e**, Social behaviours induce *c-fos* activation in the MPOA in virgin and mated males and females. Groups are labelled as follows. C, fresh bedding exposure; KO, *Trpc2*$^{-/-}$; fa, father; vf, virgin female; mo, mother. Mean + s.e.m., one-way ANOVA followed by Bonferroni's post test comparing all the social interaction groups to fresh bedding control, $***P < 0.001$. NS, not significant. **f, g**, catFISH identifying parenting and mating induced *c-fos* in the MPOA in males show that the two behaviours activate largely distinct MPOA neuronal populations. Par, Parenting; Mat, Mating; nuc, nuclear (yellow); cyto, cytoplasmic (red). Mean + s.e.m., one-way ANOVA followed by Bonferroni's post test comparing all pairs of groups, $**P < 0.01$. **h**, Co-labelling *c-fos* and *Gal* in the MPOA of virgin females after interaction with pups. **i, j**, Percentage of *c-fos*$^{+}$ cells expressing *Gal* and percentage of *Gal*$^{+}$ cells expressing *c-fos* in males and females after various social interactions, compared to the percentages of NeuN^{+} cells expressing *Gal* and *c-fos*, respectively. Agg, Aggression. Mean + s.e.m., *t*-test pairing the measurements from each animal, adjusted by Benjamini–Hochberg procedure controlling the false discovery rate. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$; NS, not significant.

Fathers and virgin females robustly activated similar brain areas after parental care, namely the anteroventral periventricular nucleus (AVPe; data not shown) and the MPOA, and these regions remained consistently silent in virgin males. Specifically, we observed striking increases in the number of MPOA *c-fos*$^{+}$ cells of maternal virgin females, *Trpc2*$^{-/-}$ virgin males and paternal fathers (Fig. 2a–e), indicating that a common pathway for parental behaviour exists in males and females that is normally repressed in virgin males by vomeronasal inputs. The ventral bed nucleus of the stria terminalis (BNST)/dorsal MPOA was shown to play an important role in rat maternal behaviour[12,22], but also in sexual behaviour[23–27], thermoregulation[28] and gonadotropin-releasing hormone (GnRH) secretion[29]. Accordingly, we observe robust MPOA *c-fos* activation after mating, medial to the area containing parenting-induced *c-fos* (Fig. 2e, f).

To determine whether parenting and mating activate different MPOA neurons, we performed a cellular compartment analysis of temporal activity by fluorescent *in situ* hybridization (catFISH)[30], allowing direct

comparison of two activated cell populations. Animals experiencing the same behaviour twice showed ~70% overlap of nuclear and cytoplasmic *c-fos* MPOA signals, whereas animals engaged in different behaviours showed only 20–30% overlap, indicating that mating and parenting activate largely distinct MPOA neuronal populations (Fig. 2f, g).

The MPOA is a highly heterogeneous structure[31], which receives inputs from, and sends information to, multiple brain regions[32,33]. The identity of cell populations governing parental behaviour is unknown. We characterized active cells in parental behaviour using double fluorescent *in situ* hybridization with *c-fos* and a series of molecular markers with distinct MPOA expression[34] (Methods). We uncovered the neuropeptide galanin (*Gal*) as a candidate marker for MPOA *c-fos*$^+$ cells in virgin females, mothers, and fathers. Across all markers surveyed, *Gal* showed the highest enrichment in parenting-induced *c-fos*$^+$ MPOA cells (Extended Data Fig. 1a, b). 38.3% ± 1.6% of MPOA *c-fos*$^+$ cells in virgin females, 43.9% ± 4.6% in mothers, and 33.4% ± 0.8% in fathers co-express *Gal* (mean ± s.e.m., *t*-test pairing each animal, $P < 0.001$ for virgin females and fathers, $P < 0.05$ for mothers; Fig. 2h, i). Further, 24.8% ± 0.8% of MPOA *Gal*$^+$ cells in females, 26.7% ± 1.4% in mothers, and 16.8% ± 0.9% in fathers co-express *c-fos* (mean ± s.e.m., paired *t*-test, $P < 0.001$ for virgin females and fathers, $P < 0.01$ for mothers; Fig. 2j). *Gal* is also found in minor subsets of mating and aggression-induced *c-fos*$^+$ cells in males, whereas overlap between *Gal* and *c-fos* induced by pup-directed aggression is not significantly different from chance level (Fig. 2i, j).

Gal is expressed in several brain areas and modulates multiple physiological functions[35]. *Gal* is also co-expressed by prolactin-secreting cells of the pituitary and involved in lactation[36]. We found that MPOA *Gal*$^+$ cell number is not sexually dimorphic, although MPOA *Gal* expression level is slightly higher in females than in males (Extended Data Fig. 1c, d). Most MPOA *c-fos*$^+$ and *Gal*$^+$ cells express *Gad1*, characteristic of GABA ergic inhibitory neurons (Extended Data Fig. 1e–h).

Ablation of MPOA *Gal*$^+$ neurons

We next investigated the requirement of MPOA *Gal*$^+$ neurons for parental behaviour in females and mated males. We obtained a Gal-Cre transgenic line (GENSAT) and confirmed appropriate *Cre* expression in MPOA *Gal*$^+$ neurons: 94.6% of the *Gal*$^+$ cells co-express *Cre* ($n = 858$ cells in 2 animals) and 94.8% of the *Cre*$^+$ cells co-express *Gal* (725 cells in 2 animals; Extended Data Fig. 2a). To specifically ablate MPOA *Gal*$^+$ neurons, Gal-Cre mice were given bilateral MPOA injections of recombinant adeno-associated virus (AAV) expressing Cre-dependent diphtheria toxin A fragment (AAV-DTA) (Extended Data Fig. 2b). On average, AAV-DTA eliminated ~60% of MPOA *Gal*$^+$ cells, compared to Gal-Cre-negative littermate controls receiving the same treatment (Extended Data Fig. 2c, d). We verified that an independent MPOA cell population expressing thyrotropin releasing hormone (*Trh*) was not affected by targeted ablation (Extended Data Fig. 2e). Furthermore, neighbouring *Gal*$^+$ cells in the AVPe, paraventricular nucleus (PVN) and dorsomedial hypothalamic nucleus (DMH) were unaffected, confirming the spatial specificity of viral-mediated ablations (Extended Data Fig. 2f–h).

Virgin females with MPOA *Gal*$^+$ neuron loss showed striking reductions in maternal behaviour and emergence of pup-directed aggression (Fig. 3) compared to Gal-Cre-negative littermates or Gal-Cre females with AAV-Flex-GFP viral injections (Extended Data Fig. 3a–f). The duration of overall maternal interaction is positively correlated with the number of remaining *Gal*$^+$ cells (Fig. 3a; $n = 23$, $P < 0.05$, $R = 0.46$). Moreover, whereas virgin females with low ablation of MPOA *Gal*$^+$ cells were maternal, females with ablation efficiencies above 50% displayed loss of maternal care with increased pup-directed aggression (Fig. 3b), accompanied by significantly reduced crouching, nest building, retrieval to nest, and maternal interaction compared to controls (Fig. 3c–h). Thus, MPOA *Gal*$^+$ cells represent an essential neuronal population for the maternal behaviour of virgin females.

Next, we examined the effects of MPOA *Gal*$^+$ cell ablation on retrieving behaviour of nursing females (Methods). Control mothers retrieved all four pups, whereas most mothers with loss of over 50% *Gal*$^+$ MPOA cells failed to retrieve pups, suggesting a critical role of *Gal*$^+$ cells in maternal behaviour of lactating females (Extended Data Fig. 4a–c).

We then tested the requirement of *Gal*$^+$ neurons for male parental behaviour (Methods). As with females, disappearance of parental behaviour in males was associated with loss of over 50% of *Gal*$^+$ cells (Fig. 4a, b).

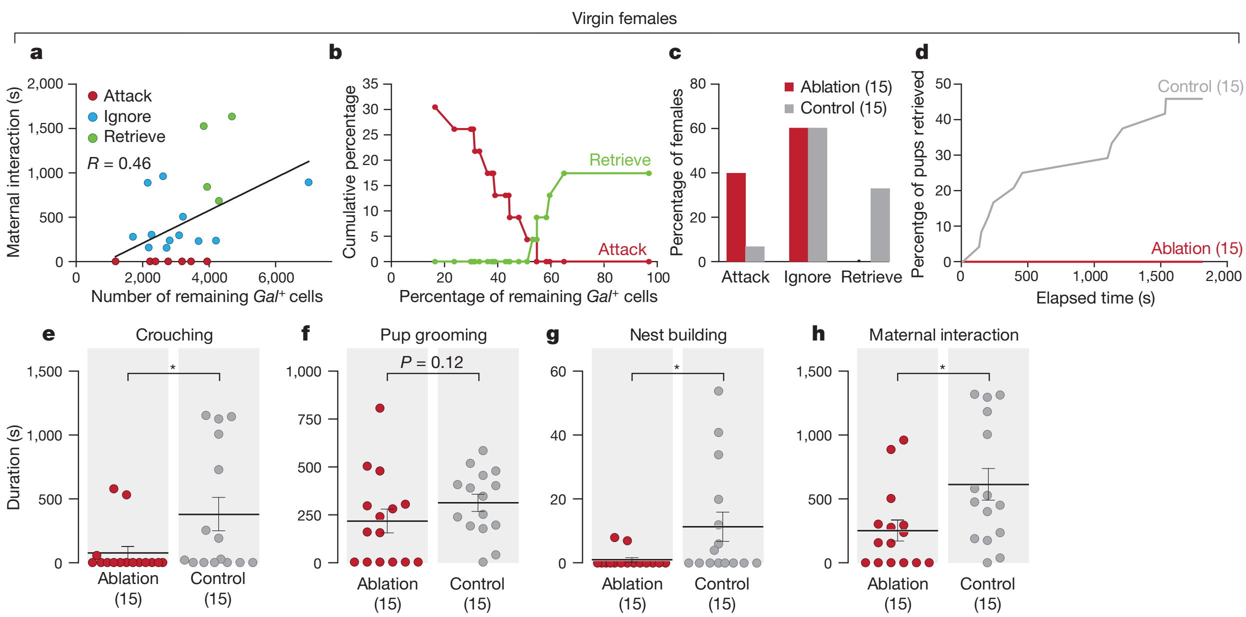


Figure 3 | Ablation of MPOA *Gal*$^+$ neurons impairs maternal behaviour in virgin females. a, Linear regression of maternal interaction and the number of remaining MPOA *Gal*$^+$ cells in ablated virgin females. Animals are colour coded by their behaviour categories. Pearson correlation, $n = 23$, $P < 0.05$, $R = 0.46$. **b**, Cumulative percentages of females that retrieved or attacked pups as a function of the percentage of remaining *Gal*$^+$ cells, $n = 23$. Reference cell number (100%) is the average MPOA *Gal*$^+$ cell number in the control group. As the remaining number of *Gal*$^+$ cells increases or decreases on the *x*-axis, each female is added to the maternal group or the infanticidal group according to its behaviour type, respectively. **c**, Behaviour of ablated females with over 50% ablation efficiency ($n = 15$) compared to control ($n = 15$). Chi-square test, $P < 0.05$. **d**, Combined percentage of pups (out of two) retrieved by the ablation group as a function of time, compared to the controls. Kolmogorov–Smirnov test, $P < 0.05$. **e–h**, Crouching (**e**), pup grooming (**f**), nest building (**g**) and maternal interaction (**h**). Mean ± s.e.m. Mann–Whitney test, *$P < 0.05$.

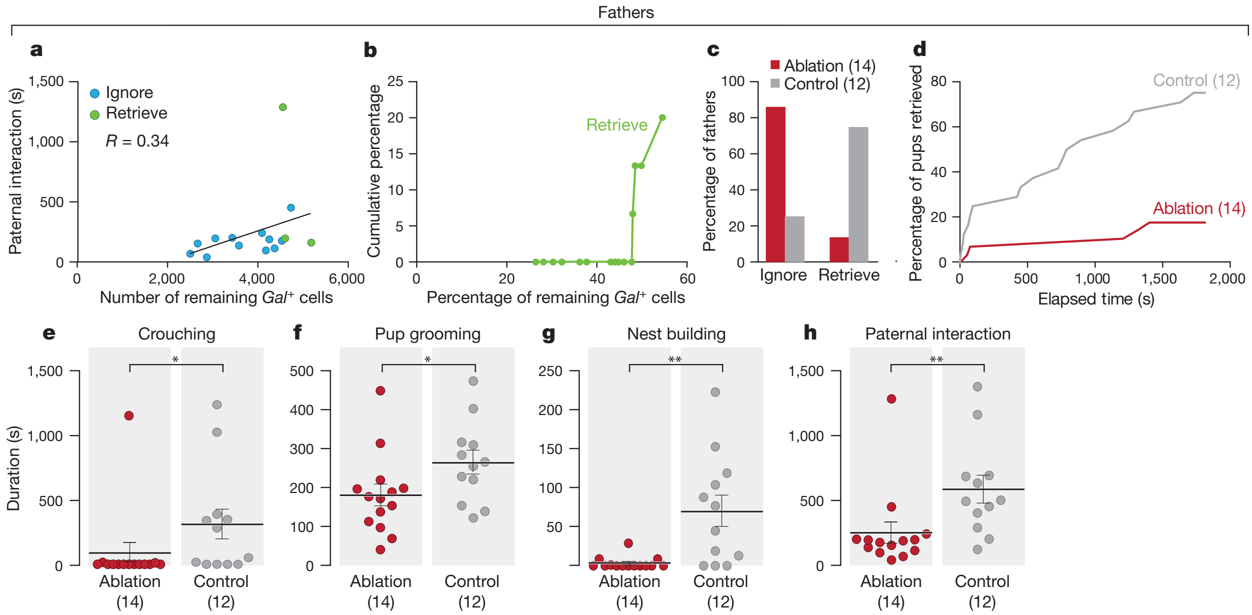


Figure 4 | Ablation of MPOA *Gal*$^+$ neurons impairs paternal behaviour in fathers. **a**, Linear regression of paternal interaction and number of remaining *Gal*$^+$ cells in the MPOA in ablated fathers. Animals are colour coded by their behaviour categories. Pearson correlation, $n = 15$, $P = 0.21$, $R = 0.34$. **b**, Cumulative percentages of paternal males (Retrieve) as a function of the percentage of remaining *Gal*$^+$ cells, $n = 15$. Reference cell number (100%) is the average MPOA *Gal*$^+$ cell number in the control group. **c**, Behaviour type of ablated fathers with over 50% ablation efficiency ($n = 14$) compared to control ($n = 12$). Fisher's exact test, $^{**}P < 0.01$. **d**, Combined percentage of pups retrieved (out of two) by the ablation group as a function of time, compared to the controls. Kolmogorov–Smirnov test, $P < 0.001$. **e**–**h**, Crouching (**e**), pup grooming (**f**), nest building (**g**) and paternal interaction (**h**). Mean ± s.e.m., Mann–Whitney test, $^{*}P < 0.05$, $^{**}P < 0.01$.

Behaviour assays showed that only 14.3% of males with over 50% MPOA *Gal*$^+$ neuronal loss ($n = 14$) displayed paternal behaviour 3 weeks after mating, compared to 75% of littermate controls ($n = 12$; Fisher's exact test, $P < 0.01$; Fig. 4c). Ablated animals showed deficits in crouching, pup grooming, nest building, retrieval to nest, and overall paternal interaction compared to controls (Fig. 4d–h).

Gal$^+$ cell ablation did not affect locomotion or inter-male aggression (Extended Data Fig. 5a–f), but decreased mounting duration and increased latency to mount (Extended Data Fig. 5g–i). This mating defect may result from ablation of the small subset of MPOA *Gal*$^+$ cells activated during mating or from interactions between brain circuits controlling parenting and mating.

To further assess the functional specificity of MPOA *Gal*$^+$ cells in behaviour control, we examined the effect of ablating MPOA tyrosine hydroxylase (*Th*) cells using AAV-DTA in *Th-IRES-Cre* males[37]. ~70% of *Th*$^+$ cells were ablated compared to littermate controls (Extended Data Fig. 6a, b). The ablation was restricted to the MPOA, as the AVPe *Th*$^+$ cells were largely unaffected (Extended Data Fig. 6c). Although MPOA *Th*$^+$ cell loss was comparable to *Gal*$^+$ cell loss (Extended Data Fig. 6d), it did not affect parenting, mating, or inter-male aggression in males (Extended Data Fig. 6e–o), highlighting the critical role of *Gal*$^+$ cells in the control of parenting.

Remarkably, specific ablation of *Gal*$^+$ cells affected all major aspects of parental behaviour. Additionally, whereas a significant fraction of virgin females with strong reduction in *Gal*$^+$ neurons attacked pups, no mated males or nursing females with high ablation efficiency displayed pup-directed aggression. This result suggests that, in virgin females, *Gal*$^+$ neurons are important for both maternal behaviour and inhibition of pup-directed aggression, whereas in fathers and mothers, mating suppresses circuits for pup-directed aggression independently of *Gal*$^+$ neuronal activation.

Activation of MPOA *Gal*$^+$ neurons

To address whether activation of MPOA *Gal*$^+$ neurons is sufficient to suppress pup-directed aggression and potentiate parental behaviour, virgin males and fathers were tested during optogenetic activation of *Gal*$^+$ neurons. *Gal-Cre* males were given MPOA-targeted injections of a Cre-dependent channelrhodopsin-2 fused with enhanced yellow fluorescent protein virus (AAV-ChR2:EYFP) and implanted with an optic fibre. Negative controls were Gal-Cre-negative littermates receiving the same treatment. In stimulation trials, blue light was delivered to the MPOA whenever the male contacted a pup with its snout. Post-mortem *in situ* hybridization confirmed specific MPOA *ChR2:EYFP* expression in *Gal*$^+$ cells (Fig. 5a, b). ~60% of MPOA *Gal*$^+$ cells expressed AAV-ChR2:EYFP, similar to the expression of AAV-DTA in ablation experiments (Extended Data Fig. 9k). Additionally, we verified that parenting-induced *c-fos*$^+$ and *c-fos*$^-$ subpopulations of *Gal*$^+$ cells showed comparable viral infection rates (Extended Data Fig. 9k). Light stimulation in awake behaving animals produced strong *c-fos* induction in MPOA *Gal*$^+$ cells of Gal::ChR2 males, but not control males (33.5% ± 3.3% for Gal::ChR2 males, 6 animals; 4.1% ± 0.2% for controls, 8 animals; mean ± s.e.m., *t*-test, $P < 0.001$).

We first investigated whether *Gal*$^+$ cell activation reduced pup-directed aggression. Each male was tested multiple times with stimulation (stim) and non-stimulation (no stim) (Methods). Light stimulation of MPOA *Gal*$^+$ neurons in Gal::ChR2 males inhibited attacking in 16 of 18 trials (6 animals, 2–4 trials per animal), whereas the same animals attacked in 18 of 19 trials without stimulation (Fig. 5c, d). Loss of pup-directed aggression was not due to pup-avoidance, as light-stimulated Gal::ChR2 virgin males displayed frequent and lengthy bouts of pup grooming not observed in controls (Fig. 5e, f and Extended Data Fig. 7). However, light stimulation did not significantly alter the behaviour of control virgin males (Fig. 5c–f and Extended Data Fig. 7).

We next observed effects of light stimulation on parental behaviour of fathers (Methods). Light stimulation elicited strikingly elevated pup grooming in Gal::ChR2 compared to non-stimulated fathers (Fig. 5g, 5i; Extended Data Fig. 8). Interestingly, induction of active pup grooming in Gal::ChR2 stimulated males was seen at the expense of crouching (Fig. 5h, i and Extended Data Fig. 8).

To address the specificity of *Gal*$^+$ cell activation in parental behaviour, we also tested other behaviours. *Gal*$^+$ cell activation left mating behaviour unaffected, but diminished inter-male aggression and increased

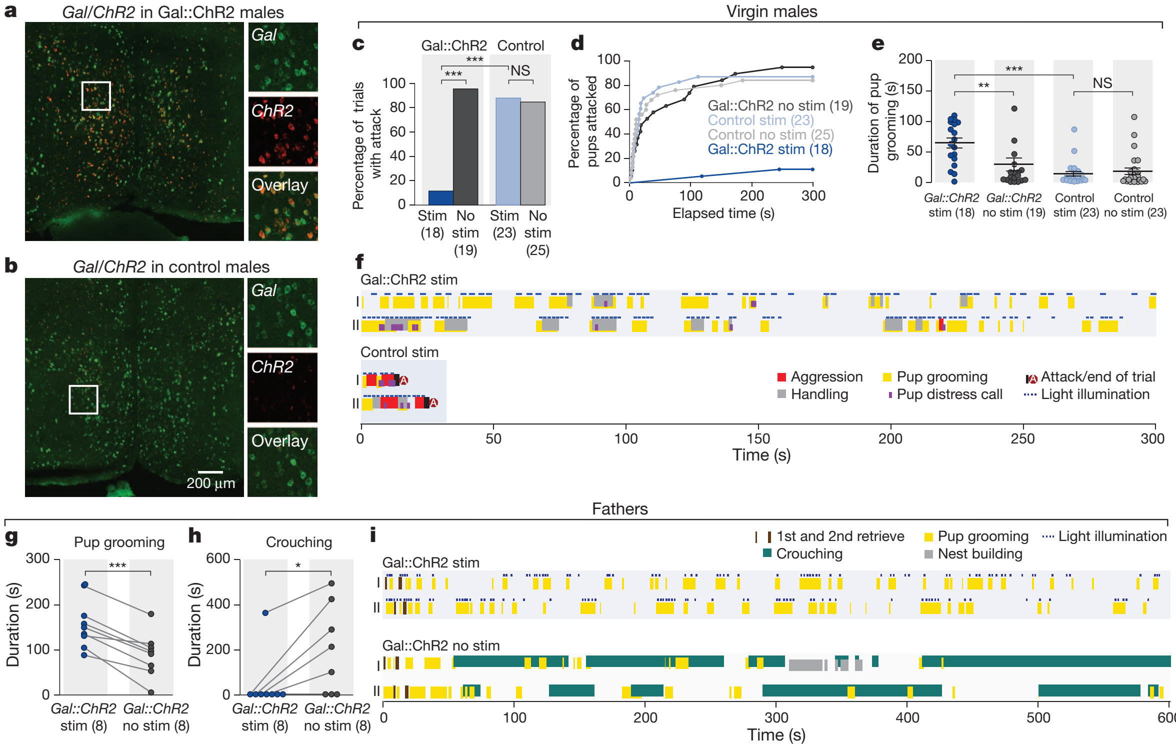


Figure 5 | Optogenetic activation of MPOA *Gal*$^+$ neurons in males suppresses attack and promotes pup grooming. **a**, **b**, Co-labelling *Gal* and *ChR2:EYFP* expression in the MPOA of Gal::ChR2 (**a**) and control males (**b**). **c**, Percentage of trials with attacks of pups by virgin males. Fisher's exact test with Bonferroni correction, $^{***}P < 0.001$; NS, not significant. **d**, Percentage of pups attacked by each group of virgin males. Gal::ChR2 stimulated (stim) trials are significantly different from Gal::ChR2 not stimulated (no stim) and control stimulated (control stim) trials. Kolmogorov–Smirnov test with Bonferroni correction, $P < 0.001$. **e**, Pup grooming in the tests with virgin males. Mean ± s.e.m.; Mann–Whitney test with Bonferroni correction. $^{**}P < 0.01$, $^{***}P < 0.001$; NS, not significant. **f**, Sample behaviour raster plot of Gal::ChR2 stimulated and control stimulated trials in virgin males. Note that two behaviour elements (such as pup grooming and handling) can occur simultaneously. **g**, Pup grooming in the tests of fathers. $n = 8$ for each group, t-test pairing the same animal with and without light stimulation, $^{***}P < 0.001$. **h**, Crouching in the tests of fathers. $n = 8$, paired t-test, $^{*}P < 0.05$. **i**, Sample behaviour raster plot of Gal::ChR2 stimulated and Gal::ChR2 not stimulated trials in tests with fathers.

locomotion (Extended Data Fig. 9a–g), whereas length of social contact was equivalent in control and stimulation trials across assays (Extended Data Fig. 9h, i). Duration of light illumination was also comparable across all stimulation experiments (Extended Data Fig. 9j).

These results indicate that optogenetic activation of MPOA *Gal*$^+$ cells is sufficient to suppress pup-directed aggression and induce active pup grooming. The suppression of inter-male aggression and increased locomotion may result from increased parenting and pup-seeking, or from other unknown behavioural drives. Surprisingly, whereas ablation of *Gal*$^+$ cells leads to mating defects, activation of these cells did not increase mating. This may reflect unknown complexity in social circuit coding, or originate from slightly different virus infectivity in ablation and activation experiments.

Discussion

Our data provide significant insights into the control of opposing social behaviours in mice: parenting versus pup-directed aggression. Whereas vomeronasal circuits in virgin males mediate aggression towards pups, this response is silenced in females and mated males, and neuronal pathways underlying parental care are activated instead. We show here that MPOA *Gal*-expressing cells are critical for the control of mouse parental behaviour and the suppression of pup-directed aggression, thus acting as a central regulatory node of social interactions with pups. Manipulation of this genetically defined neuronal population switches on or off the parental behaviour of mice, providing a precious entry point for further dissection of neural circuits underlying parental care and their modulation by social experience. The functional heterogeneity among *Gal*$^+$ cells, also reported in most neuropeptide-expressing neurons[38–40], may underlie the observed partial modulation of other social behaviours. A more refined characterization of *Gal*$^+$ neuron subpopulations may help identify subsets of MPOA neurons involved in distinct behaviours.

Interestingly, ablation of MPOA *Gal*$^+$ neurons leads to reductions in all tested aspects of parenting, whereas MPOA *Gal*$^+$ neuron activation triggers pup grooming but no other parental displays. An understanding of the natural pattern of MPOA *Gal*$^+$ neuron activity during parental interactions, particularly during intense care such as grooming versus more passive display like huddling with pups, may help optimize ChR2-mediated stimulation of MPOA *Gal*$^+$ neurons and its behavioural outcome. Additionally, although MPOA *Gal*$^+$ neuronal activity seems essential for parenting behaviour, some behavioural displays may require simultaneous activation of additional neuronal populations. Interestingly, activation of MPOA *Gal*$^+$ neurons increases locomotion without affecting social contact and decreases inter-male aggression, suggesting complex functional relationships between parenting and other behaviour circuits.

From our results, the relationship between circuits mediating parental care and pup-directed aggression is complex and modulated by social experience. Virgin males with activated MPOA *Gal*$^+$ neurons do not attack pups, indicating that these neurons suppress pup-directed aggression directly. Indeed, loss of MPOA *Gal*$^+$ neurons impairs parental behaviour and elicits pup-directed aggression in virgin females. However, MPOA *Gal*$^+$ neuron ablation suppresses parental behaviour without facilitating pup-directed aggression in mothers or fathers, suggesting that circuits underlying pup-directed aggression are silenced in mated animals through independent mechanisms. Future circuit-level analysis of MPOA *Gal*$^+$ neurons will help uncover mutual connections between circuits underlying parenting, pup-directed aggression and mating, and assess connectivity with other brain areas participating in parenting[12,41].

Finally, a variety of hormones and neuropeptides, including oestradiol, testosterone, prolactin, progesterone and oxytocin, modulate parenting according to the physiological state of the animal and its social context[42–49]. It will be interesting to determine if *Gal*, a neuropeptide involved in modulation of many homeostatic and reproductive functions is a new player in the regulation of parental behaviour.

METHODS SUMMARY

Behavioural analysis of parental behaviour was performed as described in the methods section. *In situ* hybridization and catFISH were performed as previously described. Targeted ablation was performed by injecting the MPOA bilaterally in Gal-Cre animals with AAV-DTA. For cell activation, AAV-ChR2::YFP was injected bilaterally into the MPOA of Gal-Cre animals and an optical fibre was implanted, allowing stimulation using blue light. Details of the experimental setup are provided in the Methods.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 18 May 2013; accepted 2 April 2014.

1. Numan, M. & Insel, T. R. *The Neurobiology of Parental Behavior* (Springer, 2003).
2. Lonstein, J. S. & De Vries, G. J. Sex differences in the parental behavior of rodents. *Neurosci. Biobehav. Rev.* **24,** 669–686 (2000).
3. Brown, R. Hormonal and experiential factors influencing parental behaviour in male rodents: an integrative approach. *Behav. Processes* **30,** 1–27 (1993).
4. Rosenblatt, J. S. Nonhormonal basis of maternal behavior in the rat. *Science* **156,** 1512–1514 (1967).
5. Terkel, J. & Rosenblatt, J. S. Maternal behavior induced by maternal blood plasma injected into virgin rats. *J. Comp. Physiol. Psychol.* **65,** 479–482 (1968).
6. Moltz, H., Lubin, M., Leon, M. & Numan, M. Hormonal induction of maternal behavior in the ovariectomized nulliparous rat. *Physiol. Behav.* **5,** 1373–1377 (1970).
7. Svare, B. & Mann, M. Infanticide: genetic, developmental and hormonal influences in mice. *Physiol. Behav.* **27,** 921–927 (1981).
8. Brooks, R. J. & Schwarzkopf, L. Factors affecting incidence of infanticide and discrimination of related and unrelated neonates in male *Mus musculus. Behav. Neural Biol.* **37,** 149–161 (1983).
9. Labov, J. B. Factors influencing infanticidal behavior in wild male house mice (*Mus musculus*). *Behav. Ecol. Sociobiol.* **6,** 297–303 (1980).
10. vom Saal, F. S. & Howard, L. S. The regulation of infanticide and parental behavior: implications for reproductive success in male mice. *Science* **215,** 1270–1272 (1982).
11. vom Saal, F. S. Time-contingent change in infanticide and parental behavior induced by ejaculation in male mice. *Physiol. Behav.* **34,** 7–15 (1985).
12. Numan, M. & Stolzenberg, D. S. Medial preoptic area interactions with dopamine neural systems in the control of the onset and maintenance of maternal behavior in rats. *Front. Neuroendocrinol.* **30,** 46–64 (2009).
13. Numan, M. Medial preoptic area and maternal behavior in the female rat. *J. Comp. Physiol. Psychol.* **87,** 746–759 (1974).
14. Segovia, S. & Guillamón, A. Sexual dimorphism in the vomeronasal pathway and sex differences in reproductive behaviors. *Brain Res. Brain Res. Rev.* **18,** 51–74 (1993).
15. Stowers, L., Holy, T. E., Meister, M., Dulac, C. & Koentges, G. Loss of sex discrimination and male-male aggression in mice deficient for TRP2. *Science* **295,** 1493–1500 (2002).
16. Kimchi, T., Xu, J. & Dulac, C. A functional circuit underlying male sexual behaviour in the female mouse brain. *Nature* **448,** 1009–1014 (2007).
17. Mennella, J. A. & Moltz, H. Infanticide in the male rat: the role of the vomeronasal organ. *Physiol. Behav.* **42,** 303–306 (1988).
18. Tachikawa, K. S., Yoshihara, Y. & Kuroda, K. O. Behavioral transition from attack to parenting in male mice: a crucial role of the vomeronasal system. *J. Neurosci.* **33,** 5120–5126 (2013).
19. Fleming, A., Vaccarino, F., Tambosso, L. & Chee, P. Vomeronasal and olfactory system modulation of maternal behavior in the rat. *Science* **203,** 372–374 (1979).
20. Liman, E. R., Corey, D. P. & Dulac, C. TRP2: a candidate transduction channel for mammalian pheromone sensory signaling. *Proc. Natl Acad. Sci. USA* **96,** 5791–5796 (1999).
21. Elwood, R. W. Inhibition of infanticide and onset of paternal care in male mice (*Mus musculus*). *J. Comp. Psychol.* **99,** 457–467 (1985).
22. Calamandrei, G. & Keverne, E. B. Differential expression of Fos protein in the brain of female mice dependent on pup sensory cues and maternal experience. *Behav. Neurosci.* **108,** 113–120 (1994).
23. Arendash, G. W. & Gorski, R. A. Effects of discrete lesions of the sexually dimorphic nucleus of the preoptic area or other medial preoptic regions on the sexual behavior of male rats. *Brain Res. Bull.* **10,** 147–154 (1983).
24. Dominguez, J. M. & Hull, E. M. Dopamine, the medial preoptic area, and male sexual behavior. *Physiol. Behav.* **86,** 356–368 (2005).
25. Powers, B. & Valenstein, E. Sexual receptivity: facilitation by medial preoptic lesions in female rats. *Science* **175,** 1003–1005 (1972).
26. Pfaff, D. W. & Sakuma, Y. Facilitation of the lordosis reflex of female rats from the ventromedial nucleus of the hypothalamus. *J. Physiol. (Lond.)* **288,** 189–202 (1979).
27. Bakker, J., Woodley, S. K., Kelliher, K. R. & Baum, M. J. Sexually dimorphic activation of galanin neurones in the ferret's dorsomedial preoptic area/anterior hypothalamus after mating. *J. Neuroendocrinol.* **14,** 116–125 (2002).
28. McAllen, R. M., Tanaka, M., Ootsuka, Y. & McKinley, M. J. Multiple thermoregulatory effectors with independent central controls. *Eur. J. Appl. Physiol.* **109,** 27–33 (2010).
29. Jennes, L. & Conn, P. Gonadotropin-releasing hormone and its receptors in rat brain. *Front. Neuroendocrinol.* **15,** 51–77 (1994).
30. Guzowski, J. F., McNaughton, B. L., Barnes, C. A. & Worley, P. F. Environment-specific expression of the immediate-early gene *Arc* in hippocampal neuronal ensembles. *Nature Neurosci.* **2,** 1120–1124 (1999).
31. Simerly, R. B., Gorski, R. A. & Swanson, L. W. Neurotransmitter specificity of cells and fibers in the medial preoptic nucleus: an immunohistochemical study in the rat. *J. Comp. Neurol.* **246,** 343–363 (1986).
32. Simerly, R. B. & Swanson, L. W. The organization of neural inputs to the medial preoptic nucleus of the rat. *J. Comp. Neurol.* **246,** 312–342 (1986).
33. Simerly, R. B. & Swanson, L. W. Projections of the medial preoptic nucleus: a *Phaseolus vulgaris* leucoagglutinin anterograde tract-tracing study in the rat. *J. Comp. Neurol.* **270,** 209–242 (1988).
34. Lein, E. S. *et al.* Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445,** 168–176 (2007).
35. Mechenthaler, I. Galanin and the neuroendocrine axes. *Cell. Mol. Life Sci.* **65,** 1826–1835 (2008).
36. Wynick, D. *et al.* Galanin regulates prolactin release and lactotroph proliferation. *Proc. Natl Acad. Sci. USA* **95,** 12671–12676 (1998).
37. Lindeberg, J. *et al.* Transgenic expression of Cre recombinase from the tyrosine hydroxylase locus. *Genesis* **40,** 67–73 (2004).
38. Caldwell, H. K., Lee, H.-J., Macbeth, A. H. & Young, W. S. Vasopressin: behavioral roles of an "original" neuropeptide. *Prog. Neurobiol.* **84,** 1–24 (2008).
39. Lee, H.-J., Macbeth, A. H., Pagani, J. H. & Young, W. S. Oxytocin: the great facilitator of life. *Prog. Neurobiol.* **88,** 127–151 (2009).
40. Betley, J. N., Cao, Z. F. H., Ritola, K. D. & Sternson, S. M. Parallel, redundant circuit organization for homeostatic control of feeding behavior. *Cell* **155,** 1337–1350 (2013).
41. Numan, M. *et al.* The importance of the basolateral/basomedial amygdala for goal-directed maternal responses in postpartum rats. *Behav. Brain Res.* **214,** 368–376 (2010).
42. Champagne, F. A., Diorio, J., Sharma, S. & Meaney, M. J. Naturally occurring variations in maternal behavior in the rat are associated with differences in estrogen-inducible central oxytocin receptors. *Proc. Natl Acad. Sci. USA* **98,** 12736–12741 (2001).
43. Champagne, F. A., Weaver, I. C. G., Diorio, J., Sharma, S. & Meaney, M. J. Natural variations in maternal care are associated with estrogen receptor alpha expression and estrogen sensitivity in the medial preoptic area. *Endocrinology* **144,** 4720–4724 (2003).
44. Trainor, B. C. & Marler, C. A. Testosterone, paternal behavior, and aggression in the monogamous California mouse (*Peromyscus californicus*). *Horm. Behav.* **40,** 32–42 (2001).
45. Bridges, R. S., DiBiase, R., Loundes, D. D. & Doherty, P. C. Prolactin stimulation of maternal behavior in female rats. *Science* **227,** 782–784 (1985).
46. Lucas, B. K., Ormandy, C. J., Binart, N., Bridges, R. S. & Kelly, P. A. Null mutation of the prolactin receptor gene produces a defect in maternal behavior. *Endocrinology* **139,** 4102–4107 (1998).
47. Schneider, J. S. *et al.* Progesterone receptors mediate male aggression toward infants. *Proc. Natl Acad. Sci. USA* **100,** 2951–2956 (2003).
48. Pedersen, C. A., Ascher, J. A., Monroe, Y. L. & Prange, A. J. Jr. Oxytocin induces maternal behavior in virgin female rats. *Science* **216,** 648–650 (1982).
49. Insel, T. R. & Young, L. J. The neurobiology of attachment. *Nature Rev. Neurosci.* **2,** 129–136 (2001).

Acknowledgements We thank K. Deisseroth for the Cre-dependent AAV-ChR2:EYFP construct; E. Boyden for the Cre-dependent AAV-GFP construct; N. Shah for the AAV-Flex-taCasp3-TEVp virus; S. Sullivan for behaviour annotation and scoring; R. Hellmiss for figure artwork; E. Soucy and J. Greenwood for technical assistance. We also thank members of the Dulac and Uchida laboratories and V. Murthy, A. Schier and M. Meister for advice on experiments and statistical analysis and comments on the manuscript, and the anonymous reviewers for their helpful suggestions and comments. This work was supported by the Howard Hughes Medical Institute and the National Institute of Health (NIDCD).

Author Contributions Z.W. and C.G.D. conceived and designed the study. Z.W. and A.E.A. performed the experiments and collected the data. J.F.B. and Z.W. developed the setup for ChR2-mediated cell activation. M.W.-U. constructed the AAV-DTA virus. Z.W. and C.G.D. interpreted the results and wrote the paper with comments from A.E.A., J.F.B. and M.W.-U.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.G.D. (dulac@fas.harvard.edu).

METHODS

Animals. Animals were maintained on 12 h:12 h light/dark cycle (lighted hours: 02:00–14:00) with food and water available *ad libitum*. Animal care and experiments were carried out in accordance with the NIH guidelines and approved by the Harvard University Institutional Animal Care and Use Committee (IACUC).

Trpc2 knockout mice of C57BL/6J x129/Sv mixed genetic background were generated previously in our laboratory. The complete null allele of the *Trpc2* gene locus was confirmed by western blotting[15].

The Gal-Cre BAC transgenic line (STOCK Tg(Gal-cre)KI87Gsat/Mmucd, 031060-UCD) was imported from the Mutant Mouse Regional Resource Center. In this line, a Cre recombinase cassette followed by a polyadenylation sequence is inserted at the ATG codon of the first coding exon of the *Gal* gene. The imported line was in an FVB/N-Crl:CD1(ICR) mixed genetic background and backcrossed to C56BL/6J genetic background in our breeding colony. The animals used in the study came from the F1 generation.

The Th-IRES-Cre knock-in line was imported from the European Mouse Mutant Archive (00254). An IRES-Cre construct was inserted in the 3′ untranslated end of the *Th* gene. The *Th* expression is not affected and Cre protein is produced in *Th*-expressing cells[37]. This line was generated originally in a mixed genetic background of 129/SvJ and C57BL/6J and then back crossed to C57BL/6J.

Behaviour assay. Before behaviour tests animals were housed individually for about one week. Experiments started at the beginning of the dark phase and were performed under dim red light, unless noted otherwise. Each test was videotaped (Sony DCR-HC65 camcorder in nightshot mode, Microsoft LifeCam HD-5000 or Geovision surveillance system) and the behaviours were scored by an individual blind to the genotype using the Observer 5.0 or XT 11 software (Noldus Information Technology). When one animal is tested in multiple behaviour assays, they are allowed at least 48 h rest between tests.

Parental behaviour assay of *Trpc2* knockout animals. 2- to 4-month-old, $Trpc2^{+/-}$ and $Trpc2^{-/-}$ virgin male and female littermates were individually housed for approximately one week before the test. 1- to 3-day-old naive C57BL/6J pups were used as the standard pup intruder in all the behaviour assays performed in this study. The pups are of a different strain from the $Trpc2^{-/-}$ and $Trpc2^{-/-}$ animals and therefore are not related to the resident animals. The pregnant females were separated from the stud before parturition, so the pups are not exposed to their fathers and do not carry any adult male odour. Four naive C57BL/6J pups were introduced to the home cage of each animal and placed at the farthest corner from the resident's resting nest. The first olfactory investigation marked the beginning of the assay, which then extended until 30 min after all the pups were retrieved, or until the resident attacked and wounded the pups, or for 30 min in case neither of above happened. When a pup was attacked, the assay was ended immediately and the wounded pup was euthanized.

The behaviour of the animals was categorized based on the following criterion: animals that retrieved all the pups to the nest or built a new nest around the pups within 30 min and crouched over pups were categorized as 'Retrieve'. Animals that attacked the pups within 30 min were scored as 'Attack'. All the other animals were categorized as 'Ignore'. In most of the cases, retrieving is an all-or-none event such that if an animal retrieves one pup, it retrieves all the pups. An animal is scored as 'Ignore' if it does not retrieve all four pups or does not crouch over them after retrieval. Following IACUC guidelines, behaviour assays must be stopped before animal attacking pups have the ability to kill them. Thus, to accurately describe the attack behaviour, we mainly used 'pup-directed aggression' or 'attack' instead of 'infanticide'.

The following behaviours were scored: latency to retrieve each pup (picking up a pup with its mouth and carrying it to the nesting area), latency to attack (biting a pup, often accompanied by actual wounds on the pup and confirmed immediately after the test), grooming (sniffing and licking a pup), crouching (extending its limbs, assuming a nursing-like posture and huddling over at least 2 pups), nest building (collecting and arranging nesting material and making a nest), time spent in the nest and parental interaction ('maternal interaction' for females and 'paternal interaction' for males; calculated as the cumulative time spent crouching, grooming pups, and nest-building). Grooming, crouching, time in the nest and nest building were scored as duration during the 30-min recording after all the pups were retrieved. The latencies to retrieve or attack pups were recorded in seconds. Some behavioural variability is observed in control animals across various experiments due to the different genetic background of the transgenic lines used in each experiment. $Trpc2^{+/-}$ females are in C57BL/6J × 129/Sv mixed genetic background. Gal-Cre animals were originally in FVB/N-Crl:CD1(ICR) mixed genetic background and were backcrossed to C56BL/6J in our breeding colony. Gal-Cre virgin females used in the study were from an F1 generation, and exhibited lower level of maternal behaviour than $Trpc2^{+/-}$ virgin females.

Parental behaviour assay for mated males (Fig. 1g). $Trpc2^{+/-}$ virgin males were individually housed and then paired with females, which were checked daily for vaginal plugs in the next few days. Once a plug was spotted, the day was marked as day 0 for the mating pair and that pair was randomly assigned to a group for different length of cohabitation (1–2 days, 10–12 days, 17–20 days or 25–27 days). According to their group, the males were tested one day after the females and their litters (if any) were removed from their home cage. For example, animals tested on day 1 were separated from their mates on day 0. The animal tested on day 20 was separated from its mate on day 19 and was not exposed to its own litter. The negative controls for this essay were individually housed $Trpc2^{+/-}$ virgin males.

Mating behaviour assay. ~8 weeks old, receptive virgin females (as determined by vaginal smear) of C57BL/6J background were introduced to the resident mouse cage. Each test runs for 15 min and was videotaped and scored for the following parameters: sniffing, mounting and mounting with pelvic thrust.

Inter-male aggression assay. ~8 weeks old, castrated male of C57BL/6J background (castration performed by the Jackson Laboratory) swabbed with 50 µl fresh urine from intact wild-type males were introduced to the resident mouse cage. Every 15 min test was videotaped and scored for the following parameters: attack, sniffing and grooming intruder.

Open field test. Animals are tested for 5 min in a 60 cm × 60 cm square open arena under normal lighting. The position of the animals is tracked and analysed by Ethovision XT 8 software to calculate the distance moved, average velocity and the time spent in the centre zone. The centre zone is defined as the centre square (42 cm × 42 cm) which comprises 50% of the total area.

RNA *in situ* hybridization. Fresh brain tissues were collected from animals housed in their home cage or 35 min after the start of the behaviour tests when *c-fos* expression is analysed. For social behaviour induced *c-fos* analysis, the behaviour paradigm is generally as described in the Behaviour assay section. Only animals that actually displayed a certain behaviour were selected, that is, males that displayed mounting behaviour or females that were mounted were selected for mating induced *c-fos* analysis, males that attacked intruder for inter-male aggression induced *c-fos* analysis, animals that crouched over pups in a nest for parenting induced *c-fos* analysis, and males that attacked pups for *c-fos* induced by pup-directed aggression. The dissected brains were embedded in OCT (Tissue-Tek) and frozen with dry ice. 20-µm cryosections were used for mRNA *in situ* hybridization. Adjacent sections from each brain were usually collected over a few replicate slides to generate copies for staining with multiple probes.

Fluorescent mRNA *in situ* hybridization was performed largely as described[50]. Complementary DNA of *c-fos*, *Gal*, *Trh*, *Th*, *Gad1*, *Vglut2*, *EYFP*, *GFP*, *ChR2*, *Cre*, *mCherry* mRNA and other MPOA molecular markers (*Esr1*, *Esr2*, *Cyp19a1*, *Ar*, *Pgr*, *Prlr*, *Hcrt*, *Cart*, *Tac1*, *Penk*, *Bdnf*, *Peg10*, *Pvalb*, *Calb1*, *Calb2*, *Vip*, *Nos1*, *Cck*, *Sst*, *Nts*, *NR5a1*, *Npy*) were cloned in approximately 800-base-pair (whenever possible) segments into pCRII-TOPO vector (Invitrogen). Antisense complementary RNA (cRNA) probes were synthesized with T7 or Sp6 polymerases (Promega) and labelled with digoxigenin (DIG; Roche), fluorescein (FITC; Roche) or dinitrophenol (DNP; PerkinElmer). Where necessary and possible, a cocktail of 2–4 probes were generated covering different segments of the target mRNA to maximize strength of signal.

mRNA hybridization was performed with 0.5–1.0 ng µl^{-1} cRNA probes at 68 °C. The probes were detected using horseradish peroxidase (POD)-conjugated antibodies (anti-FITC-POD at 1/250 dilution, Roche; anti-DIG-POD at 1/500 dilution, Roche; anti-DNP-POD at 1/100 dilution, PerkinElmer). The signals were amplified using biotin-conjugated tyramide (PerkinElmer) and subsequently visualized with Alexa Fluor 488-conjugated streptavidin or Alexa Fluor 568-conjugated streptavidin (Invitrogen), or directly visualized with TSA plus cyanine 3 system, TSA plus cyanine 5 system or TSA plus Fluorescein system (PerkinElmer). Tissues were mounted with Vectashield (Vector labs) containing 8 µg ml^{-1} 4′,6-diamidino-2-phenylindole.

For catFISH, animals were subject to two 5-min episodes of behaviours interleaved with a 30 min interval, and were euthanized immediately after the second episode. The *c-fos* cytoplasmic signal induced by the first behaviour episode was compared to the *c-fos* nuclear signal induced by the second, allowing direct comparison of the two activated cell populations. The same cRNA *c-fos* probes described above were used to detect cytoplasmic signal as well as nuclear signal, and an intron probe[51] containing the first intron of the *c-fos* gene was used to detect only the nuclear signal.

Immunohistochemistry. Immunohistochemistry was performed according to standard protocols. NeuN was detected with primary antibody Mouse Anti-NeuN (1:3,000; Millipore, MAB377) and then amplified by Alexa Fluor 555 donkey anti-mouse IgG (1:500; Life Technologies).

Image analysis and cell counting. All the microscopy images were acquired with AxioImager Z2 and AxioVision software with a ×10 objective (Zeiss). Brain areas were determined based on landmark structures and white matters such as the ventricles, anterior commissure and optic tract, with the occasional assistance of Nissl staining and other area-specific molecular markers on adjacent sections when necessary. Areas of interest in the *c-fos* expression analysis included the MPOA, anteroventral

periventricular nucleus, bed nucleus of stria terminalis, medial amygdala, posteromedial cortical amygdala, nucleus accumbens, lateral septal nucleus, suprachiasmatic nucleus, paraventricular nucleus, anterior basomedial nucleus, ventromedial hypothalamic nucleus and dorsomedial hypothalamic nucleus. After manual assignment of brain structures, automated cell counting was performed using ImageJ with custom-written macro scripts. Sample images were manually counted by experimenters blind to the test condition to verify the reliability of automated cell counting. For a given brain area, the absolute cell number was determined by summing up the cell counts of all the sections deemed as part of that area, adjusted by the number of the slicing replicates collected in cryosectioning.

Targeted cell ablation in the MPOA. The rAAV8/EF1α-mCherry-Flex-dtA (AAV-DTA) construct was generated using the A subunit of the diphtheria toxin gene from a PGKdtabpA plasmid (Addgene plasmid 13440)[52]. The recombinant vectors were then serotyped with AAV8 coat proteins and packaged by the viral vector core at the University of North Carolina. AAV-DTA (4×10^{12} viral particles ml^{-1}) was injected bilaterally in the MPOA of Gal-Cre or Th-IRES-Cre males in the amount of 0.8 μl on each side (Bregma: 0.0 mm, midline: +0.5 mm; dorsal surface: −5.0 mm) with Nanoject II injector (Drummond Scientific). The negative control for Gal^+ cell ablation consisted of Cre- littermates receiving the same treatment. In the cell ablation of nursing mothers, one animal injected with AAV-Flex-taCasp3-TEVp[53] (3×10^{12} viral particles ml^{-1}) to achieve better ablation efficiency was included in the data.

The AAV-CAG-Flex-GFP (AAV-GFP) construct was developed by E. Boyden and it was packaged in serotype 8 by viral vector core at the University of North Carolina. AAV8-GFP (6×10^{12} viral particles ml^{-1}) was injected in the same manner as described above in Gal-Cre^+ animals as controls for Gal^+ and Th^+ cell ablation. It was also used to assess the infection rate of the MPOA Gal^+ and parenting-induced $c\text{-}fos^+$ cells, since AAV-DTA infection leads to cell death and prevents an accurate estimation. To test the infection rates, Gal-Cre females with AAV-GFP injections were subject to a standard parental assay and then analysed by *Gal/c-fos/GFP* triple mRNA *in situ* hybridization.

For parental behaviour, virgin females were allowed about 4 weeks of recovery, enabling optimal DTA expression and cell ablation before behaviour testing. Each female was individually housed and tested with two C57BL/6 pups, in a similar manner as described earlier. Retrieving, attacking, crouching, pup grooming, nest building and overall maternal interaction were scored. For parental behaviour test of the fathers, males were allowed about one week of recovery after surgery and then paired with females until the females gave birth (~3 weeks). 1–2 days after the pups were born, males were separated from their mates and litters, individually housed for 2–3 days and tested in a 30-min behaviour assay with two C57BL/6J pups. Retrieving, attacking, crouching, pup grooming, nest building and overall paternal interaction were scored. For mothers, females were allowed about one week of recovery after injection and then paired with males, which were removed from the females about 1 week before term. On P0, after removing the litters from a mother, 4 of the pups were re-introduced into the cage and retrieving behaviour was observed for 10 min. The brains were collected after behaviour assays for histological analysis.

ChR2-mediated cell activation. The AAV-EF1α-DIO-hChR2(H134R):EYFP (AAV-ChR2:EYFP) construct was a gift of K. Deisseroth[54] and the recombinant AAV vectors were serotyped with AAV5 coat proteins and packaged by the viral vector core at the University of North Carolina. Gal-Cre males were tested with pups and those attacked pups were selected for surgery. 0.8 μl of AAV-ChR2 (4×10^{12} viral particles ml^{-1}) was injected bilaterally into the MPOA of Gal-Cre males (Bregma: 0.0 mm, midline: +0.5 mm; dorsal surface: −5.0 mm) using Nanoject II injector (Drummond Scientific). After injection, a small plastic adaptor holding an optical fibre (300-μm diameter; Polymicro technologies) was implanted above the MPOA and affixed to the skull with dental cement (Bregma: 0.0 mm, midline: +0.2 mm; dorsal surface: −4.2 mm). The implant was positioned close to the midline to cover the MPOA in both hemispheres and lowered to a depth of approximately 0.8 mm above the centre of the AAV injection. A threaded plastic cap (Plastics One) was used to cover the implant during recovery and between experiment sessions. Gal-Cre-negative males treated with the same procedure were the negative controls.

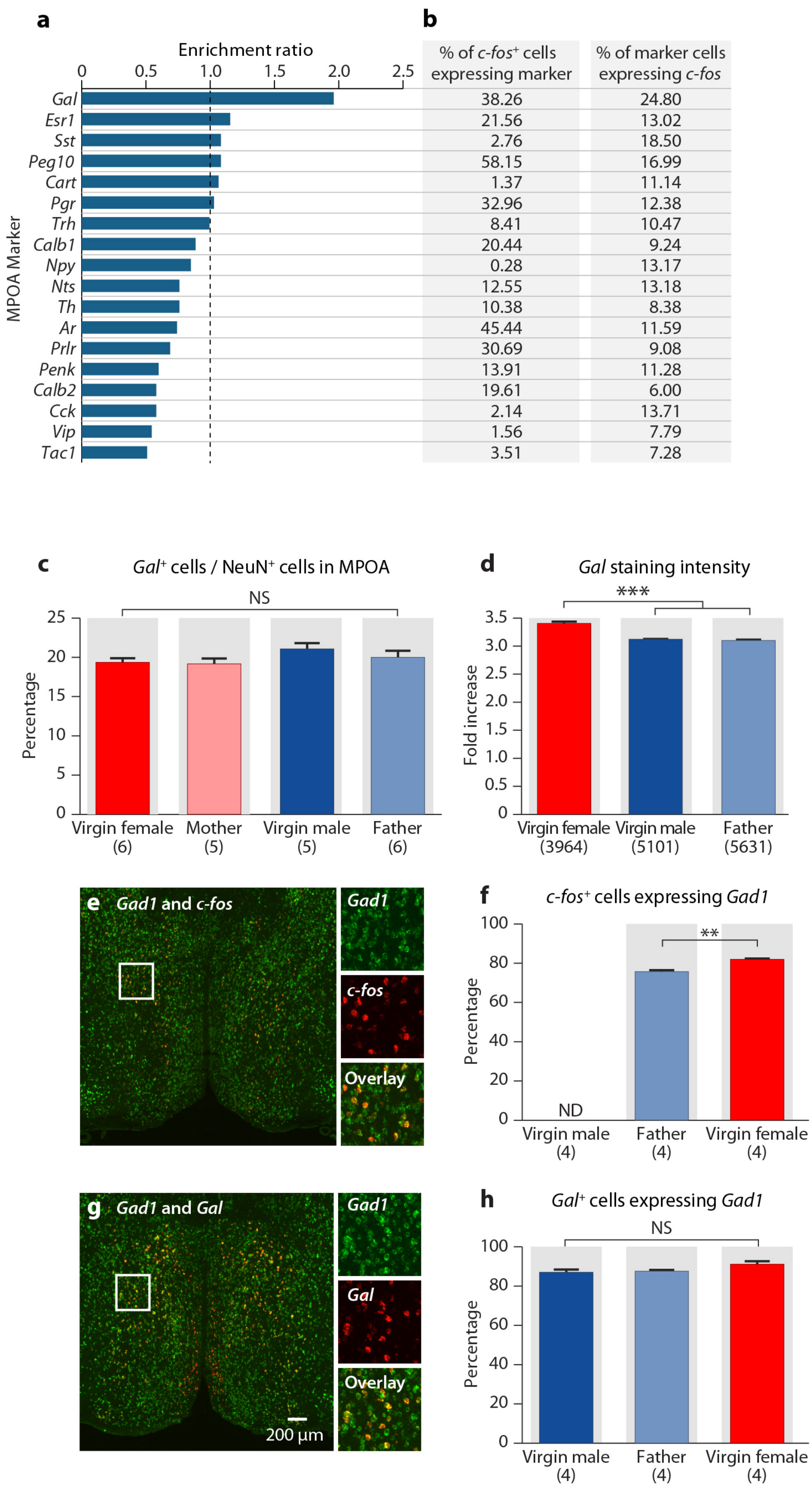
The males were tested after at least 2 weeks of recovery. Before stimulation, the implant was connected to an optical fibre (300-μm diameter, Polymicro technologies), which was connected in turn to a blue laser via an optical commutator permitting free movement of the animals. The optic fibre was flexible and long enough to allow the animal to freely behave and interact with the intruder. Both Gal::ChR2 and control animals were tested for 2–4 trials with stimulation (stim) and non-stimulation (no stim) trials randomly assigned in 1:1 ratio. In each trial, one C57BL/6J pup was introduced to the male's home cage to minimize the number of pups used in this assay, as most of the males are likely to attack pups. Blue light (473 nm) was delivered in 30-ms pulses at 20 Hz for 1–4 s whenever the male contacted the pup with its snout. The light power exiting the fibre tip was at ~10–20 mW, ensuring a light intensity above ~1.0 mW mm^{-2} over the entire MPOA[55]. There was almost no leakage of light from the optic fibre or the adaptor. Each trial was up to 5 min but when the male attacked and wounded the pup, the trial was ended and the pup was euthanized immediately. The following behaviour was scored and quantified: pup grooming (as the male sniffs or licks the pup), handling (as the male holds the pup with two forepaws), aggression (as the male grabs the pup violently and attempts to bite, usually does not wound the pups but cause them to struggle and make distress calls) and pup distress calls (only audible calls were recorded).

For paternal behaviour assays, the Gal::ChR2 and the control males were paired with females. After their pups were born, the females and the pups were removed and the males were tested in their home cage by introducing two C57BL/6J pups. Each male was tested in two 10-min trials with one stimulation and one non-stimulation trial in randomized order. Blue light is delivered when the males sniff or lick the pups. None of the males attacked pups or displayed obvious aggression. Retrieving, pup grooming, crouching and nest building behaviours were scored and quantified as described above.

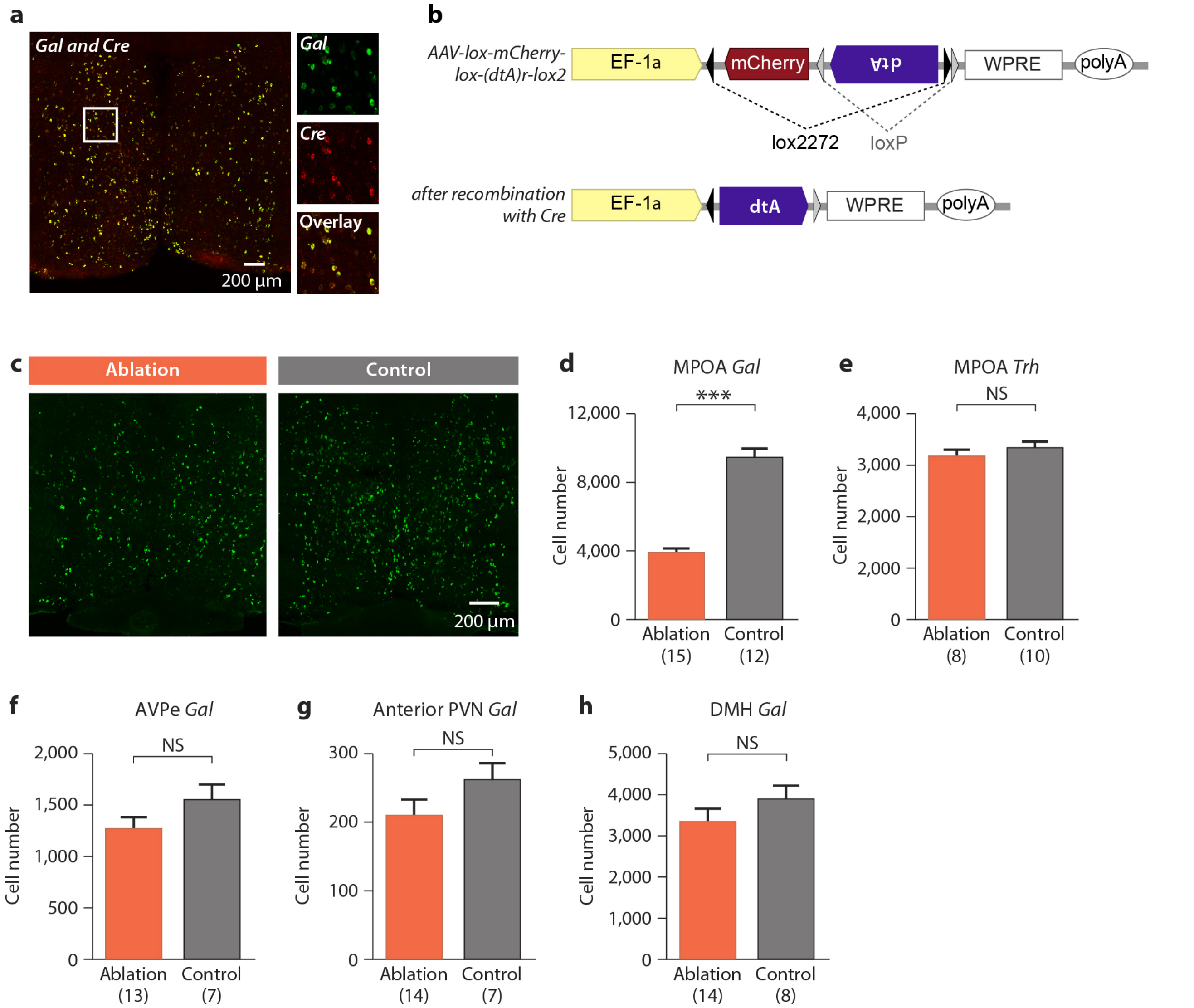
After behaviour assays, the brain tissues of these animals were collected after a standard *c-fos* induction protocol to analyse the efficiency of viral infection and cell activation. A train of light was delivered in 30-ms pulses at 20 Hz for 2 s, repeated every 10 s for 15 min, at experimental light intensity. Co-labelling between *Gal*, ChR2: EFYP and *c-fos* was analysed by mRNA *in situ* hybridization. Two Gal::ChR2 animals with less than 20% of MPOA Gal^+ cells expressing *c-fos* were discarded from the group. The fibre implants from both Gal::ChR2 and control animals were verified for efficient light transmission.

Statistics. The sample sizes in our study were chosen based on common practice in animal behaviour experiments. Data were first tested with Lilliefors test for normality. If the null hypothesis that the data come from a normal distribution cannot be rejected, Student's t-test was used. Otherwise, the Mann–Whitney test was used. Due to the strong non-normality of the behaviour data, Mann–Whitney test was used for all the behaviour analysis. For categorical data, Fisher's exact test was used.

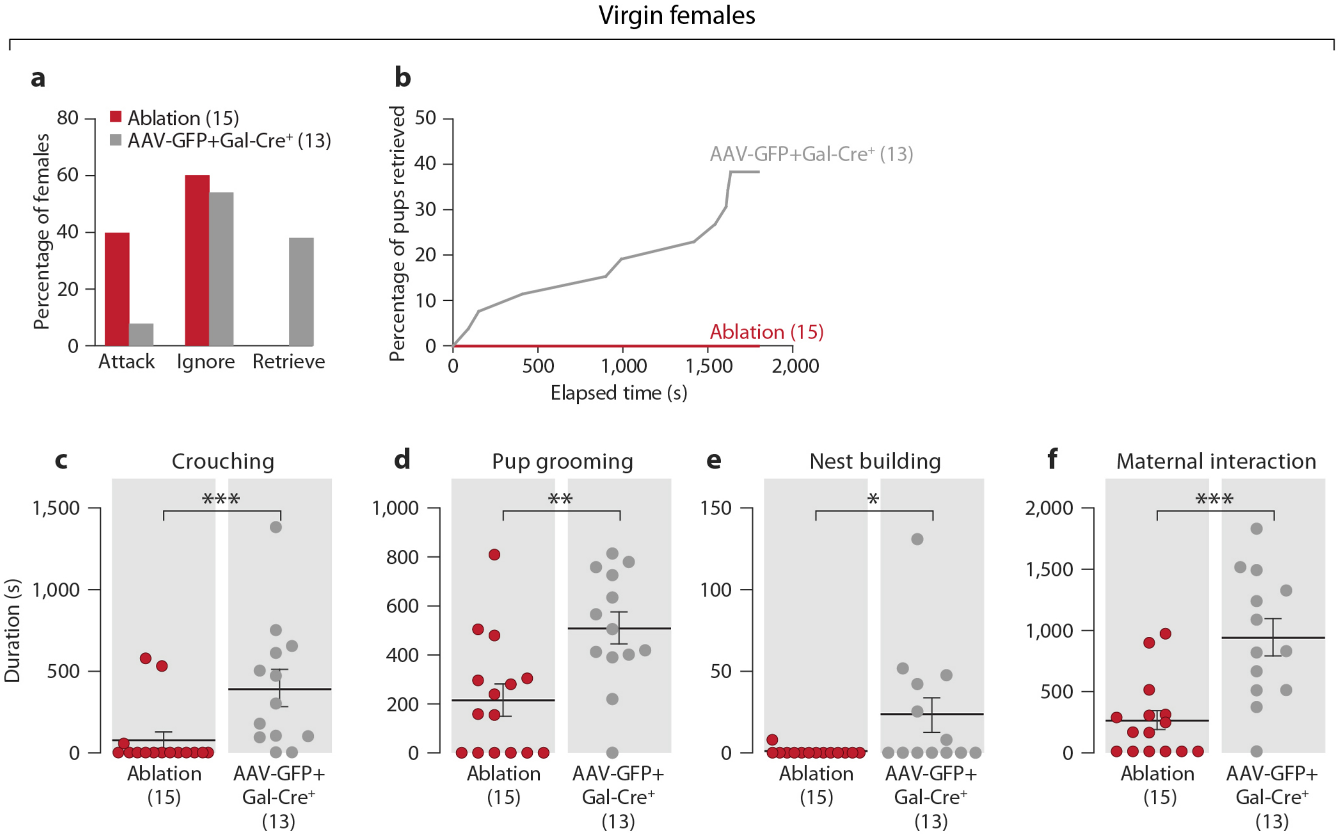
50. Isogai, Y. *et al.* Molecular organization of vomeronasal chemoreception. *Nature* **478,** 241–245 (2011).
51. Lin, D. *et al.* Functional identification of an aggression locus in the mouse hypothalamus. *Nature* **470,** 221–226 (2011).
52. Soriano, P. The PDGFα receptor is required for neural crest cell development and for normal patterning of the somites. *Development* **124,** 2691–2700 (1997).
53. Yang, C. F. *et al.* Sexually dimorphic neurons in the ventromedial hypothalamus govern mating in both sexes and aggression in males. *Cell* **153,** 896–909 (2013).
54. Gradinaru, V., Mogri, M., Thompson, K. R., Henderson, J. M. & Deisseroth, K. Optical deconstruction of parkinsonian neural circuitry. *Science* **324,** 354–359 (2009).
55. Yizhar, O., Fenno, L. & Davidson, T. Optogenetics in neural systems. *Neuron* **71,** 9–34 (2011).



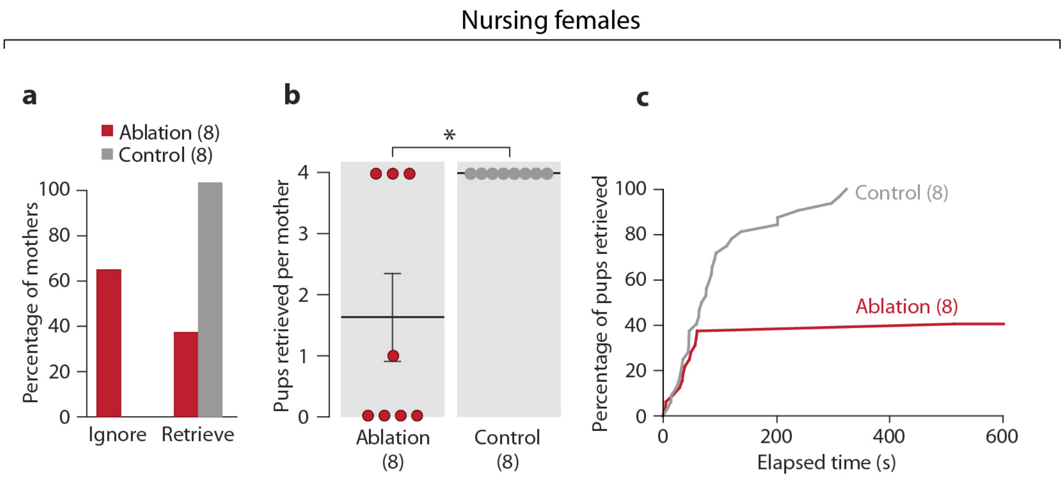
Extended Data Figure 1 | Identification of *Gal* as a marker for cells involved in parenting and characterization of MPOA *Gal*$^+$ cells. **a**, Enrichment ratio of markers in parenting induced MPOA *c-fos* in virgin females. The enrichment ratio of a given marker is calculated as the percentage of the *c-fos*$^+$ cells co-expressing the marker, divided by the percentage of NeuN$^+$ cells co-expressing this marker. **b**, The percentages of parenting induced MPOA *c-fos*$^+$ cells co-expressing markers and the percentages of marker cells co-expressing *c-fos*. **c**, Percentages of *Gal*$^+$ cells in the MPOA in virgin and sexually experienced males and females fail to identify any sexual dimorphism in MPOA *Gal*$^+$ cell representation. Mean + s.e.m., one-way ANOVA, $P > 0.2$. **d**, Fold increase of *Gal* mRNA *in situ* staining intensity compared to background in virgin females, virgin males and fathers. *Gal* mRNA expression is slightly higher (10% increase) in females than in males. Mean + s.e.m., one-way ANOVA, $^{***}P < 0.001$, NS, not significant. **e**, **f**, Percentages of *c-fos*$^+$ cells co-expressing *Gad1* in fathers and virgin females. ND, not determined. Mean + s.e.m., *t*-test, $^{**}P < 0.01$. **g**, **h**, Percentages of *Gal*$^+$ cells co-expressing *Gad1* in virgin males, fathers and virgin females. Mean + s.e.m., one-way ANOVA, $P > 0.1$.



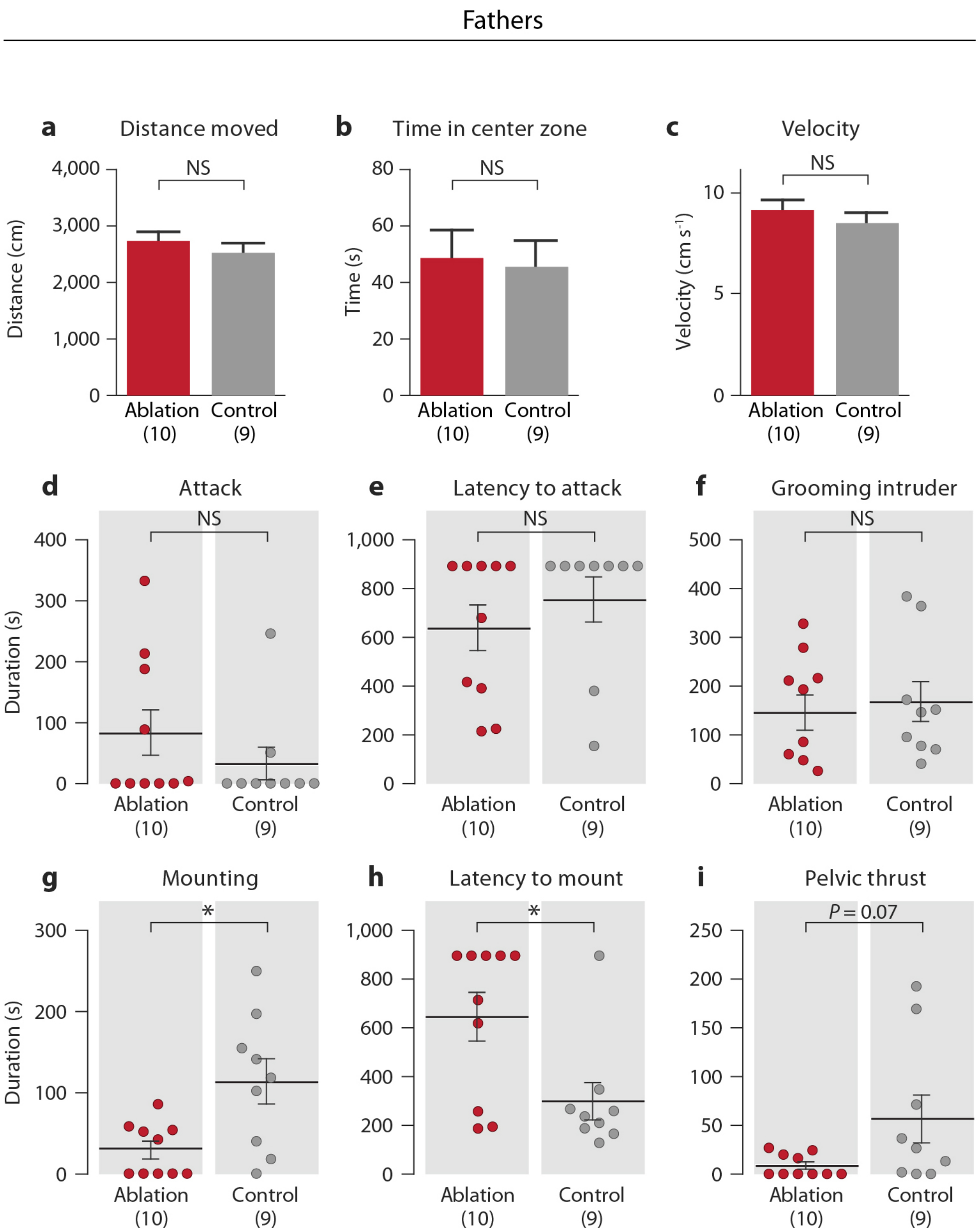
Extended Data Figure 2 | Targeted *Gal*$^{+}$ cell ablation in the MPOA. **a**, Co-labelling of *Gal* and *Cre* expressing cells by mRNA *in situ* hybridization in Gal-Cre females indicates near perfect overlap. **b**, Schematic map of the Cre-dependent AAV-DTA virus; DTA is doubly flanked by two sets of incompatible lox sites and inverted to enable transcription after Cre-mediated recombination. **c**, *Gal* mRNA expression in the MPOA of ablated and control males. **d**, Number of MPOA *Gal*$^{+}$ cells in ablation group compared to controls. Mean + s.e.m., *t*-test, $^{***}P < 0.001$. **e**, Number of MPOA *Trh*$^{+}$ cells in the ablation group and control. Mean + s.e.m., *t*-test, $P > 0.2$. **f–h**, *Gal*$^{+}$ cell numbers in the AVPe (**f**), anterior part of the PVN (**g**) and the DMH (**h**) in MPOA targeted ablation compared to control. Mean + s.e.m., *t*-test, $P > 0.1$.



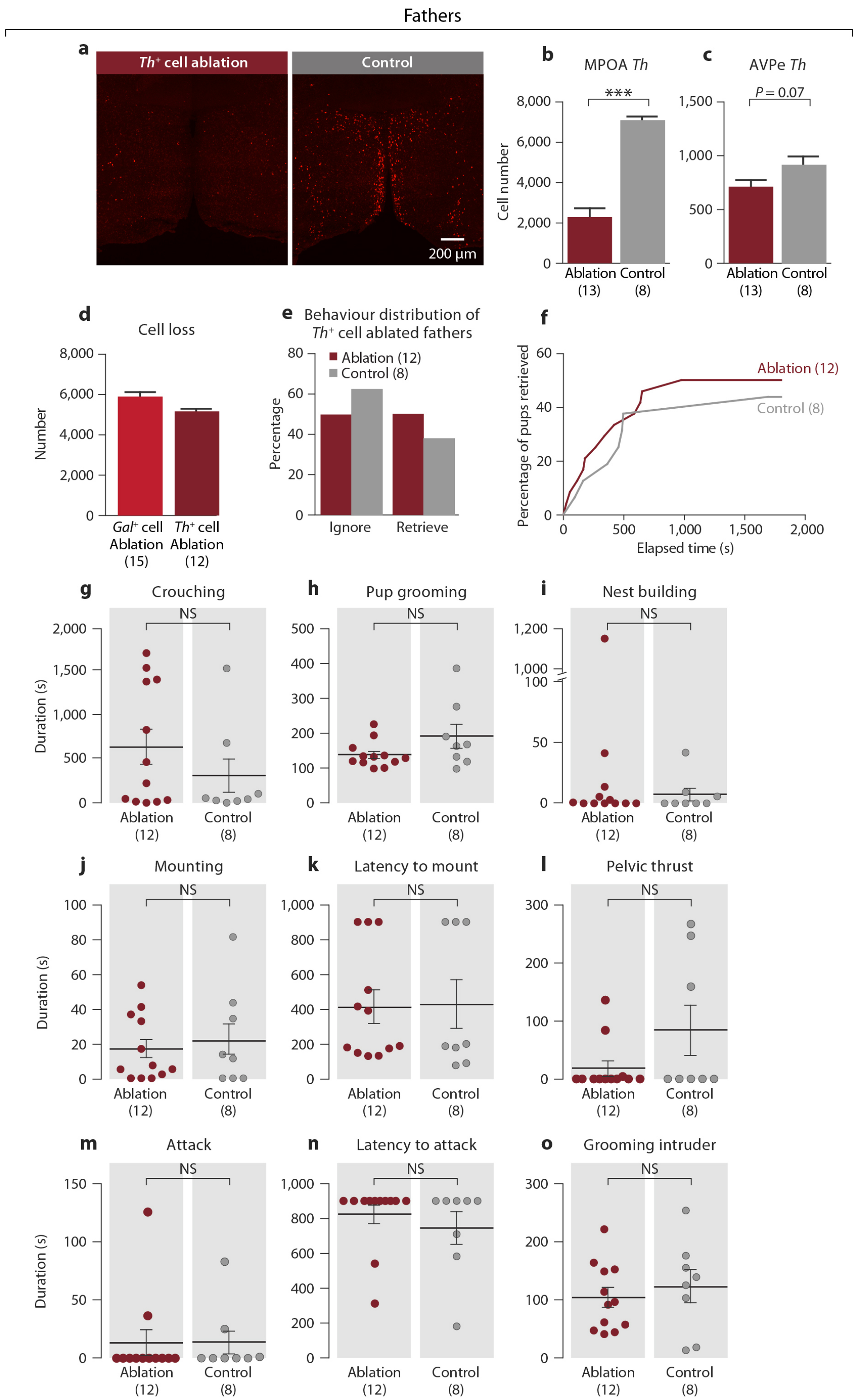
Extended Data Figure 3 | Females with MPOA Gal^+ cell ablation compared to Gal-Cre$^+$ controls injected with AAV-Flex-GFP. **a**, Behaviour of MPOA Gal^+ cell ablated virgin females with over 50% ablation efficiency ($n = 15$) compared to Gal-Cre$^+$ controls injected with AAV-Flex-GFP ($n = 13$). Chi-square test, $P < 0.05$. **b**, Percentage of pups retrieved by Gal^+ cell ablated virgin females as a function of time compared to the controls. The retrieving data of the two pups in each test are combined. Kolmogorov–Smirnov test, $P < 0.05$. **c–f**, Crouching (**c**), pup grooming (**d**), nest building (**e**) and maternal interaction (**f**) in the Gal^+ cell ablated virgin females and control. Mean ± s.e.m. Mann–Whitney test, $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$. The control females with the longest crouching and of nest building duration are different individuals.



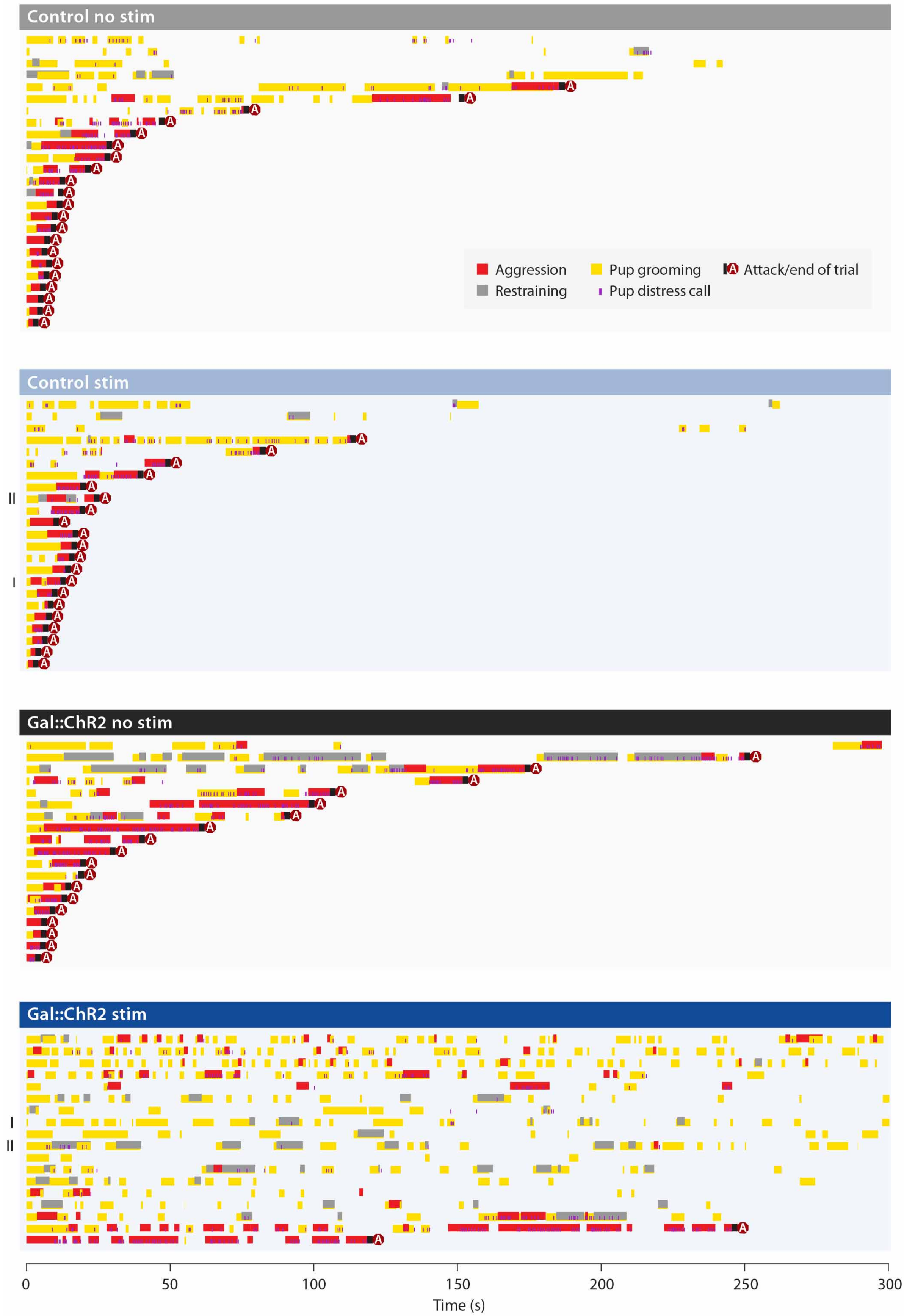
Extended Data Figure 4 | Deficits in retrieving behaviour of mothers with MPOA *Gal*$^{+}$ cell ablation. **a**, Behaviour of MPOA *Gal*$^{+}$ cell ablated mothers ($n = 8$) compared to controls ($n = 8$). Fisher's exact test, $P < 0.05$. **b**, Number of pups retrieved by each mother. Mean ± s.e.m. Mann–Whitney test, $^{*}P < 0.05$. **c**, Percentage of pups retrieved by the ablation group as a function of time compared to the controls. The retrieving data of the four pups in each test are combined. Kolmogorov–Smirnov test, $P < 0.001$.



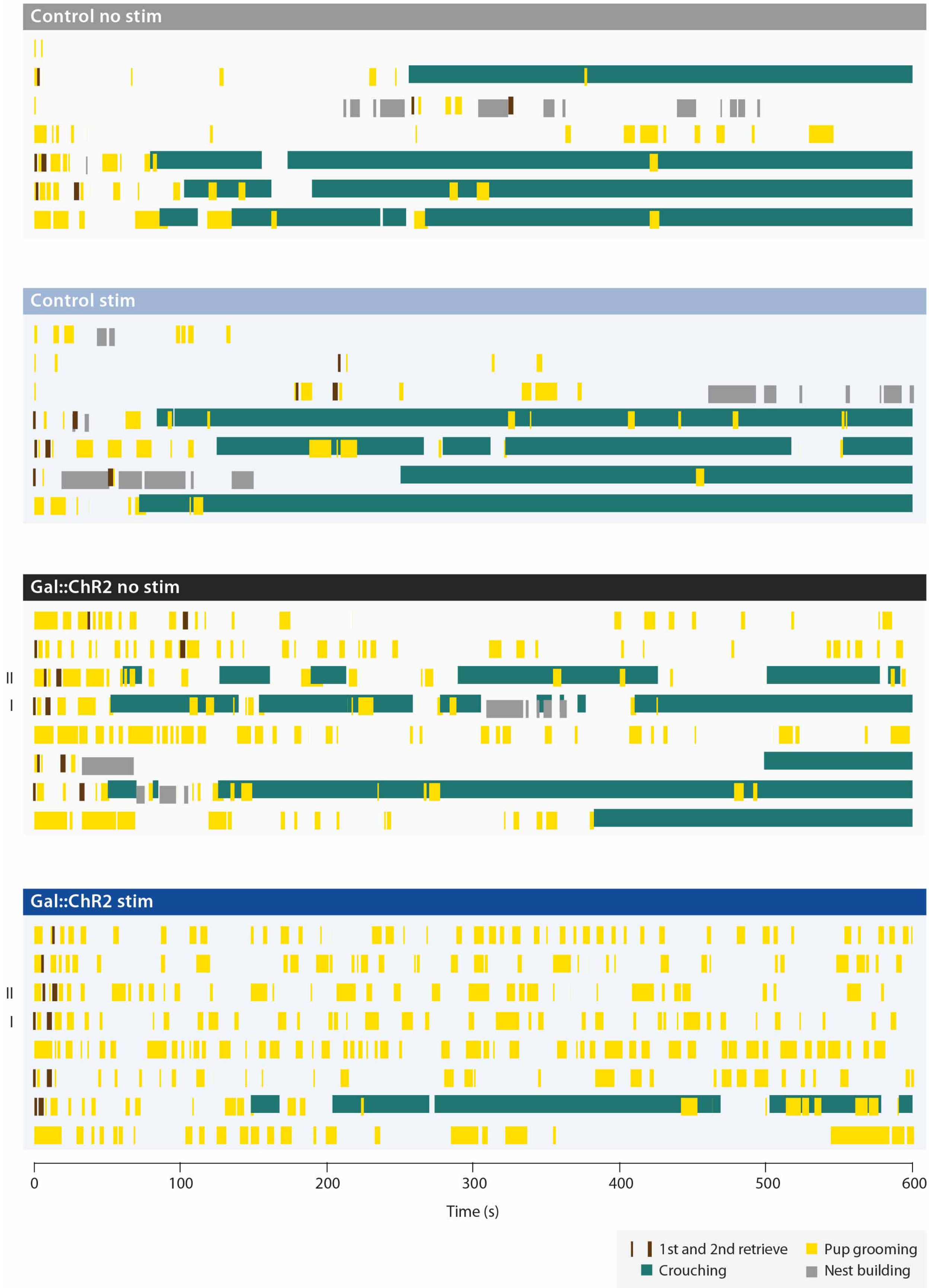
Extended Data Figure 5 | Mating, inter-male aggression and locomotor activity of MPOA *Gal*$^+$ cell ablated fathers. **a-c**, Locomotor behaviour of MPOA *Gal*$^+$ cell ablated and control fathers in a 5 min test in an open arena, measuring the distance moved (**a**), time spent in the centre zone (**b**) and the average velocity (**c**). Mean + s.e.m., *t*-test, $P > 0.3$. **d–f**, Inter-male aggression of MPOA *Gal*$^+$ cell ablated and control fathers, measuring duration of attack (**d**), latency to attack (**e**) and duration of grooming the intruder (**f**). Mean ± s.e.m. Mann–Whitney test, $P > 0.2$. **g–i**, Duration of mounting (**g**), latency to mount (**h**) and duration of mounting with pelvic thrust (**i**) of MPOA *Gal*$^+$ cell ablated fathers compared to controls. Mean ± s.e.m. Mann–Whitney test, $^*P < 0.05$.



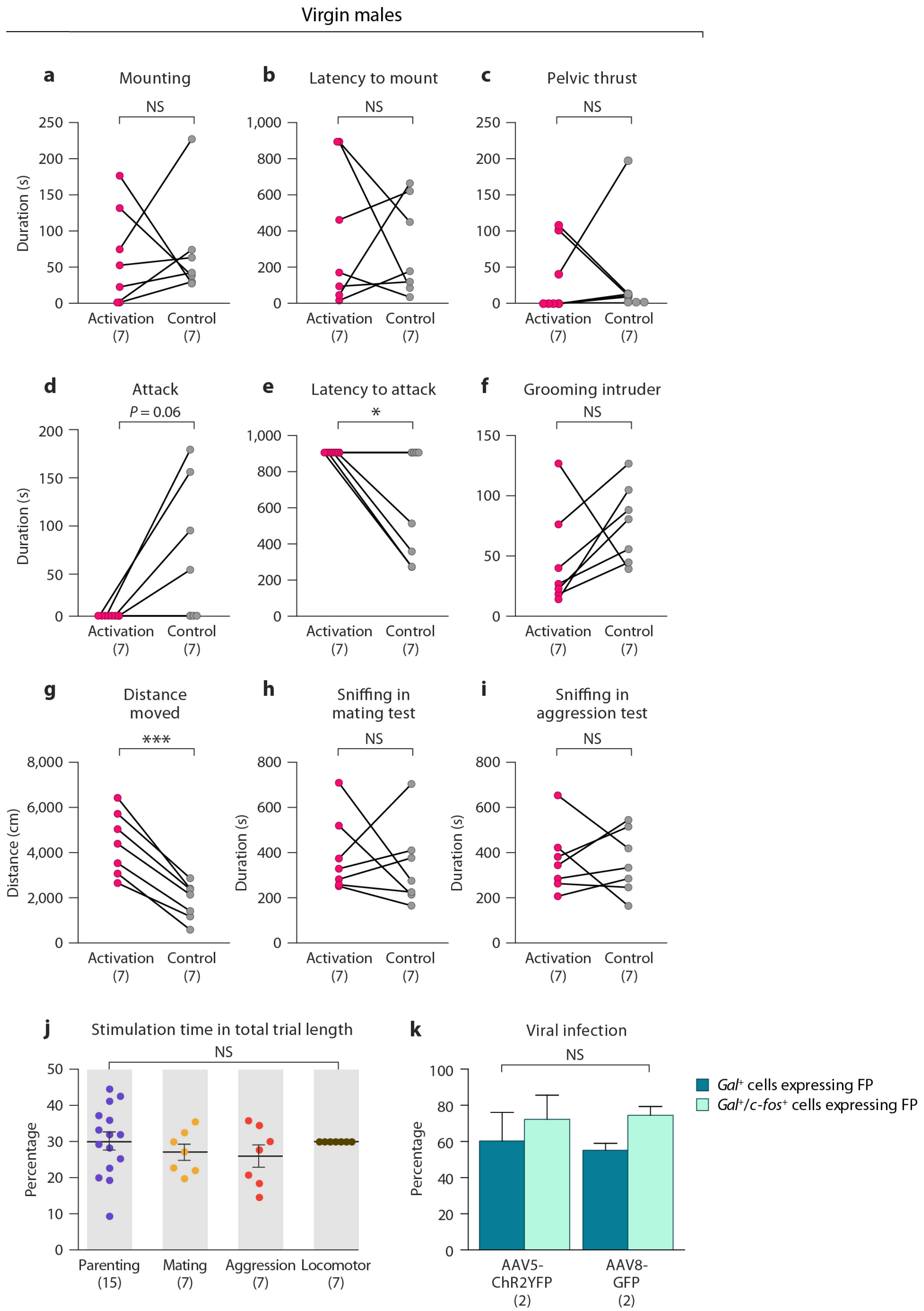
Extended Data Figure 6 | Parenting, mating and inter-male aggression of MPOA *Th*$^+$ cell ablated fathers. **a**, *Th* mRNA expression in the MPOA of *Th*$^+$ cell ablated and control fathers. **b**, Number of MPOA *Th*$^+$ cells in ablation group compared to controls. Mean + s.e.m., *t*-test, ***$P < 0.001$. **c**, Number of AVPe *Th*$^+$ cells in MPOA targeted ablation. Mean + s.e.m., *t*-test, $P = 0.07$. **d**, The number of MPOA *Th*$^+$ cell loss compared to the *Gal*$^+$ cell ablation experiments. One male had a failed *Th*$^+$ cell ablation and was removed from the data set hereafter. The *Th*$^+$ cell loss is ~87% of the *Gal*$^+$ cell loss. **e**, Behaviour type of MPOA *Th*$^+$ cell ablated fathers compared to controls. Fisher's exact test, $P > 0.6$. **f**, Combined percentage of pups (out of two) retrieved by the *Th*$^+$ cell ablation group as a function of time compared to the controls. Kolmogorov–Smirnov test, $P > 0.9$. **g–i**, Crouching (**g**), pup grooming (**h**) and nest building (**i**) in the *Th*$^+$ cell ablated fathers and control. Mean ± s.e.m. Mann–Whitney test, $P > 0.2$. The control male with the longest pup grooming also has the longest nest building activity, but not the longest duration of crouching. **j–l**, Duration of mounting (**j**), latency to mount (**k**) and duration of mounting with pelvic thrust (**l**) of MPOA *Th*$^+$ cell ablated males compared to control in a mating assay. Mean ± s.e.m. Mann–Whitney test, $P > 0.3$. **m–o**, Duration of attack (**m**), latency to attack (**n**) and duration of grooming the intruder (**o**) in MPOA *Th*$^+$ cell ablated males compared to control in an inter-male aggression assay. Mean ± s.e.m. Mann–Whitney test, $P > 0.3$.



Extended Data Figure 7 | Behaviour raster plot of Gal::ChR2 and control virgin males with and without light illumination. Each row represents a single trial lasting for 5 min or until the male attacked the pup. Trials are grouped by experiment conditions and sorted by trial length. Roman numerals indicate the sample trials shown in Fig. 5f. Various elements of the behaviour are colour coded and labelled in the insert.



Extended Data Figure 8 | Behaviour raster plot of mated Gal::ChR2 and control males with and without light illumination. Each row represents a 10-min trial. Trials are grouped by experiment conditions. Roman numerals indicate the sample trials shown in Fig. 5i. Various elements of the behaviour are colour coded and labelled in the insert.



Extended Data Figure 9 | Mating, inter-male aggression and locomotor activity of virgin males with MPOA *Gal*$^+$ cell activation and controls of light stimulation and viral infection. **a–c**, Duration of mounting (**a**), latency to mount (**b**) and duration of mounting with pelvic thrust (**c**) in virgin males with *Gal*$^+$ cell activation compared to controls in a mating assay. Paired *t*-test, $P > 0.7$. **d–f**, Duration of attack (**d**), latency to attack (**e**) and duration of grooming the intruder (**f**) in virgin males with *Gal*$^+$ cell activation compared to controls in an inter-male aggression assay. Paired *t*-test, $*P < 0.05$, NS. not significant. **g**, Distance moved in virgin males with *Gal*$^+$ cell activation compared to controls. Paired *t*-test, $***P < 0.001$. **h**, **i**, Time spent sniffing the intruder in mating (**h**) and inter-male aggression (**i**) assay. Paired *t*-test, $P > 0.6$. **j**, The duration of light stimulation in each behaviour test as a percentage of the total trial length. Mean + s.e.m., one-way ANOVA, $P > 0.6$. **k**, The percentages of *Gal*$^+$ and *Gal*$^+$/*c-fos*$^+$ cells co-expressing fluorescent protein, in females injected with AAV5-Flex-ChR2-EYFP or AAV8-Flex-GFP after maternal interaction with pups. Mean + s.e.m., two-way ANOVA examining the differences in the infection of the two viruses and the two cell populations, $P > 0.2$ for both factors and the interaction between them.

ARTICLE

doi:10.1038/nature13240

Space–time wiring specificity supports direction selectivity in the retina

Jinseop S. Kim[1]*, Matthew J. Greene[1]*, Aleksandar Zlateski[2], Kisuk Lee[1], Mark Richardson[1]†, Srinivas C. Turaga[1]†, Michael Purcaro[1], Matthew Balkam[1], Amy Robinson[1], Bardia F. Behabadi[3], Michael Campos[3], Winfried Denk[4], H. Sebastian Seung[1]† & the EyeWirers[5]

How does the mammalian retina detect motion? This classic problem in visual neuroscience has remained unsolved for 50 years. In search of clues, here we reconstruct Off-type starburst amacrine cells (SACs) and bipolar cells (BCs) in serial electron microscopic images with help from EyeWire, an online community of 'citizen neuroscientists'. On the basis of quantitative analyses of contact area and branch depth in the retina, we find evidence that one BC type prefers to wire with a SAC dendrite near the SAC soma, whereas another BC type prefers to wire far from the soma. The near type is known to lag the far type in time of visual response. A mathematical model shows how such 'space–time wiring specificity' could endow SAC dendrites with receptive fields that are oriented in space–time and therefore respond selectively to stimuli that move in the outward direction from the soma.

Compared to cognitive functions such as language, the visual detection of motion may seem trivial, yet the underlying neural mechanisms have remained elusive for half a century[1,2]. Some retinal outputs (ganglion cells) respond selectively to visual stimuli moving in particular directions, whereas retinal inputs (photoreceptors) lack direction selectivity (DS). How does DS emerge from the microcircuitry connecting inputs to outputs?

Research on this question has converged upon the SAC (Fig. 1a, b). A SAC dendrite is more strongly activated by motion outward from the cell body to the tip of the dendrite, than by motion in the opposite direction[3]. Therefore a SAC dendrite exhibits DS, and outward motion is said to be its 'preferred direction'. Note that it is incorrect to assign a single such direction to a SAC, because each of the cell's dendrites has its own preferred direction (Fig. 1a). DS persists after blocking inhibitory synaptic transmission[4], when the only remaining inputs to SACs are BCs, which are excitatory. As the SAC exhibits DS but its BC inputs exhibit little or none[5], DS appears to emerge from the BC–SAC circuit.

Mouse BCs have been classified into multiple types[6], with different time lags in visual response[7,8]. Motion is a spatiotemporal phenomenon: an object at one location appears somewhere else after a time delay. Accordingly, DS might arise because different locations on the SAC dendrite are wired to BC types with different time lags. More specifically, we propose that the proximal BCs (wired near the SAC soma) lag the distal BCs (wired far from the soma).

Such 'space–time wiring specificity' could lead to DS as follows (Fig. 1c). Motion outward from the soma will activate the proximal BCs followed by the distal BCs. If the stimulus speed is appropriate for the time lag, signals from both BC groups will reach the SAC dendrite simultaneously, summing to produce a large depolarization. For motion inward towards the soma, BC signals will reach the SAC dendrite asynchronously, causing only small depolarizations. Therefore the dendrite will 'prefer' outward motion, as observed experimentally[3].

Three-dimensional reconstruction by crowd and machine

We tested our hypothesis by reconstructing Off BC–SAC circuitry using e2198, an existing data set of mouse retinal images from serial block-face scanning electron microscopy (SBEM)[9]. The e2198 data set was oversegmented by an artificial intelligence into groups of neighbouring voxels that were subsets of individual neurons. These 'supervoxels' were assembled by humans into accurate three-dimensional (3D) reconstructions of neurons. For this activity, we hired and trained a small number of workers in the laboratory, and also transformed work into play by mobilizing volunteers through EyeWire, a website that turns 3D reconstruction of neurons into a game of colouring serial electron microscopy images.

Through EyeWire, we wanted to enable anyone, anywhere, to participate in our research. The approach is potentially scalable to extremely large numbers of 'citizen scientists'[10]. More importantly, the 3D reconstruction of neurons requires highly developed visuospatial abilities, and we wondered whether a game could be more effective[11] than traditional methods of recruiting and creating experts.

In gameplay mode, EyeWire shows a 2D slice through a 'cube', an e2198 subvolume of $256 \times 256 \times 256$ greyscale voxels (Fig. 2a). Gameplay consists of two activities: colouring the image near a location, or searching for a new location to colour. Colouring is done by clicking at any location in the 2D slice, which causes the supervoxel containing that location to turn blue. Searching is done by translating and orienting the slice within the cube, and interacting with a 3D rendering of the coloured supervoxels.

When the player first receives a cube, it already comes with a 'seed', a contiguous set of coloured supervoxels. The challenge is to colour all the rest of the supervoxels that belong to the same neuron, and avoid colouring other neurons. Gameplay for a cube terminates when the player clicks 'submit', receives a numerical score (Extended Data Fig. 1a), and proceeds to the next cube. Because our artificial intelligence is sufficiently accurate, colouring supervoxels is faster than manually colouring voxels, an older approach to 3D reconstruction[12].

[1]Brain & Cognitive Sciences Department, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [2]Electrical Engineering and Computer Science Department, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [3]Qualcomm Research, 5775 Morehouse Drive, San Diego, California 92121, USA. [4]Max-Planck Institute for Medical Research, D-69120 Heidelberg, Germany. [5]https://eyewire.org. †Present addresses: 601 N 42nd Street, Seattle, Washington 98103, USA (M.R.); Princeton Neuroscience Institute and Computer Science Deptartment, Princeton, New Jersey 08544, USA (H.S.S.); Gatsby Computational Neuroscience Unit, London WC1N 3AR, UK (S.C.T.).
*These authors contributed equally to this work.

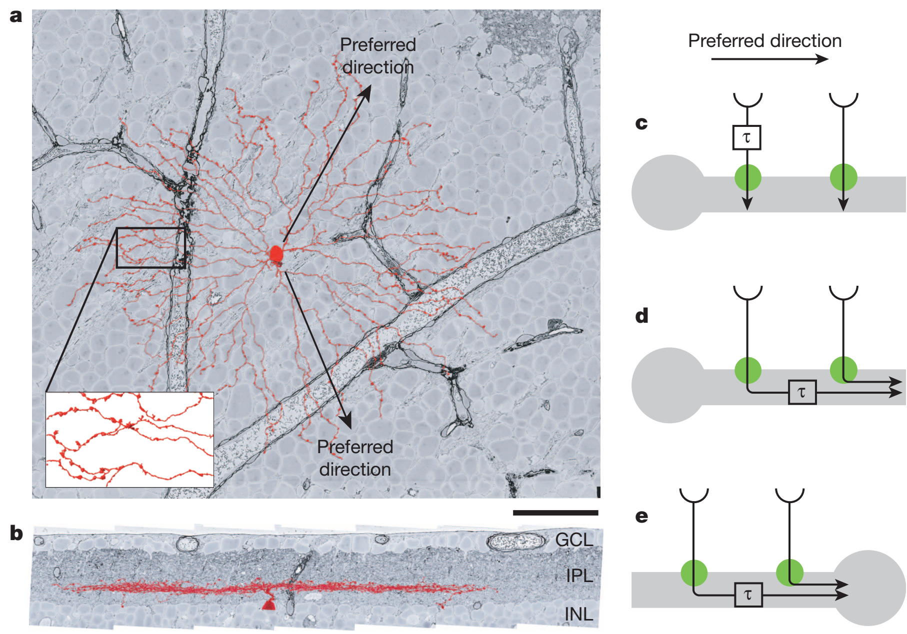


Figure 1 | Starburst amacrine cell and its direction selectivity. **a**, **b**, Off SAC (red) viewed opposite (**a**) and perpendicular (**b**) to the light axis. GCL, ganglion cell layer. Greyscale images from the e2198 data set[9]. Swellings of distal dendrites are presynaptic boutons (inset). Scale bar, 50 µm. **c**, We propose that a SAC dendrite is wired to pathways with time lags of visual response that differ by an amount τ. **d**, A previous model invoked the time lag due to signal conduction in a passive dendrite[24]. **e**, The previous model predicts an inward preferred direction for the somatic voltage, contrary to empirical observations[3].

The scoring system is designed to reward accurate colouring. This is nontrivial because EyeWire does not know the correct colouring. Each cube is assigned to multiple players (typically 5 to 10), and high scores are earned by players who colour supervoxels that other players also colour. In other words, the scoring system rewards agreement between players, which tends to be the same as rewarding accuracy.

Consensus is used not only to incentivize individual players, but also to enhance the accuracy of the entire system. Any player's colouring is equivalent to a set of supervoxels. Given the colourings of multiple players starting from the same seed in the same cube, a consensus can be computed by voting on each supervoxel. EyeWirer consensus was much more accurate than any individual EyeWirer (Fig. 2b, c).

Colouring a neuron is more challenging than it sounds. Images are corrupted by noise and other artefacts. Neurites take paths that are difficult to predict, and can branch without warning. Careless errors result from lapses in attention. Extensive practice is required to achieve accuracy. The most accurate EyeWirers (Fig. 2c, top right corner) often had experience with thousands of cubes. Improvements in accuracy were observed over the course of hundreds of cubes, corresponding to tens of hours of practice (Fig. 2d). According to subjective reports of EyeWirers, learning continues for much longer than that. By contrast, previous successes at 'crowdsourcing' image analysis involved tasks that did not require such extensive training[10,13].

Reconstructing an entire neuron requires tracing its branches through thousands of cubes. This process is coordinated by an automatic spawner, which inspects each consensus cube for branches that exit the cube. Each exit generates a new cube and seed, which are added to a queue. EyeWirers are automatically assigned to cubes by an algorithm that attempts to balance the number of plays for each cube.

Over 100,000 registered EyeWirers have been recruited by news reports, social media and the EyeWire blog. Players span a broad range of ages and educational levels, come from over 130 countries, and the

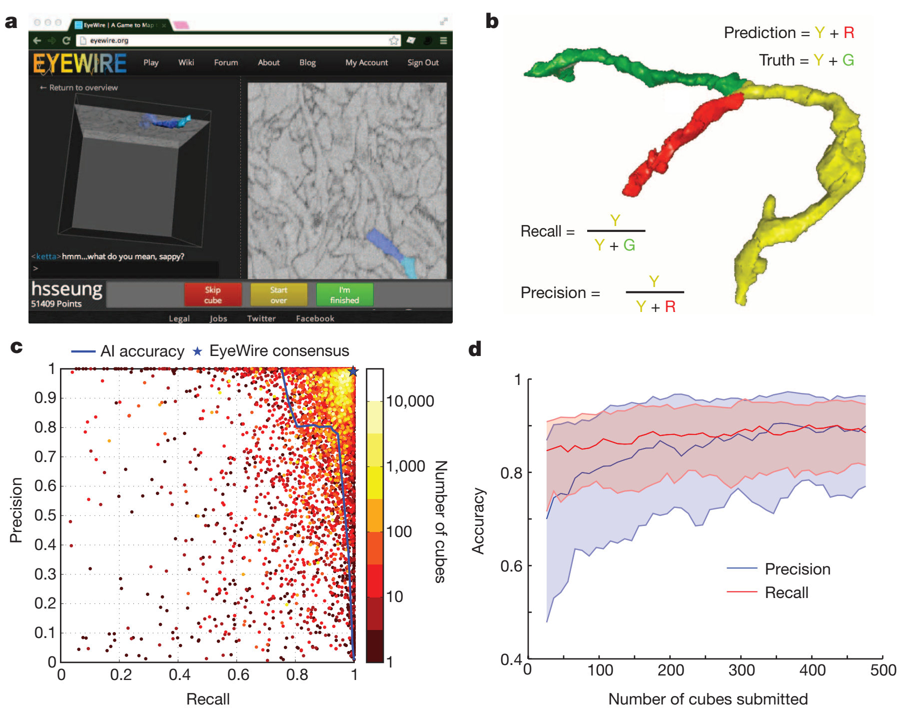


Figure 2 | EyeWire combines crowd and artificial intelligence. **a**, 3D and 2D views in the neuron reconstruction game. **b**, Precision and recall are two measures of accuracy. **c**, Accuracy of artificial intelligence (AI), 5,881 EyeWirers, and EyeWirer consensus on reconstruction of a ganglion cell. **d**, EyeWirer precision and recall increase with number of cubes submitted. Solid lines are median values across 208 EyeWirers who submitted at least 500 cubes, and shaded regions indicate 25th to 75th percentile.

great majority have no formal training in neuroscience (Extended Data Figs 2 and 3 and Supplementary Notes). These statistics show that EyeWire indeed widens participation in neuroscience research. At the same time, the most avid players constitute an elite group with disproportionate achievements. For example, the top 100 players have contributed about half of all cubes completed in EyeWire.

Laboratory workers also reconstructed neurons independently of EyeWire, with a more sophisticated version of the user interface (Methods). Their reconstructions were pooled with those of EyeWirers for the analyses reported below. Reconstruction error was quantified (Methods), and was treated like other kinds of experimental error when calculating confidence intervals from our data.

Contact analysis

We reconstructed 195 Off BC axons and 79 Off SACs from e2198 (Fig. 3b and Extended Data Fig. 4). The e2198 retina was stained in an unconventional way that did not mark intracellular structures such as neurotransmitter vesicles[9], and reliable morphological criteria for identification of BC presynaptic terminals are unknown. As an indirect measure of connectivity, contact areas were computed for all BC–SAC pairs. The resulting 'contact matrix' was analysed through two subsequent steps.

In the first step, Off BC axons were classified into five cell types, following structural criteria[14] established to correspond with previous molecular definitions[6] (Methods and Extended Data Fig. 5). BC types stratify at characteristic depths in the inner plexiform layer (IPL), and vary in size (Fig. 4a). The BCs of each type formed a 'mosaic', meaning that cells were spaced roughly periodically (Extended Data Fig. 6a–e). This is generally accepted as an important defining property of a retinal cell type. Type densities (Extended Data Fig. 6f) were roughly consistent with previous reports[6]. When the columns of the contact matrix were sorted by BC type (Fig. 4b), it became evident that BC2 and BC3a contact SACs more than other BC types.

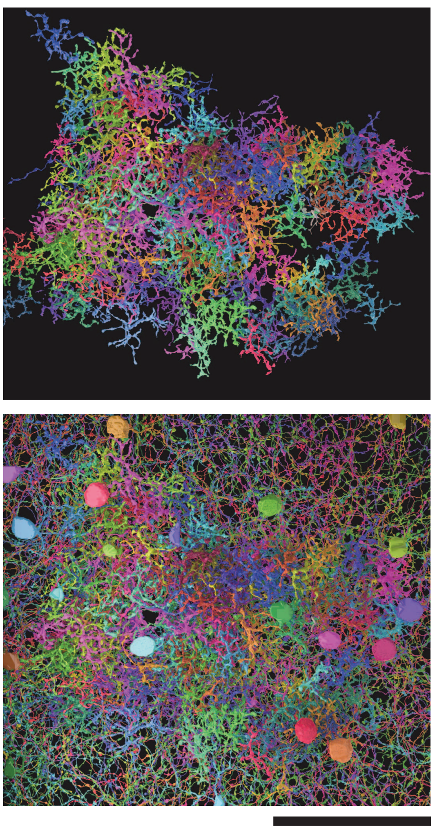


Figure 3 | 3D reconstructions of Off BCs and SACs. **a**, **b**, Cells viewed opposite the light axis. BCs alone (**a**); BCs with SACs (**b**). Scale bar, 50 μm.

In the second step, we averaged contact area over BC–SAC pairs of the same BC type and similar distance between the BC axon and the SAC soma in the plane tangential to the retina (Fig. 4c). These absolute areas were normalized to convert them into the percentage of SAC surface area covered by BCs of a given type (Methods). The resulting graphs show that BC2 prefers to contact SAC dendrites close to the SAC soma, whereas BC3a prefers to contact far from the soma (Fig. 4d and Extended Data Fig. 7c).

Imaging of intracellular calcium in BC axons[7] and extracellular glutamate around BC axons[8] indicates that BC2 lags BC3a in visual responses by 50–100 ms. Therefore BC–SAC wiring appears to possess the space–time specificity appropriate for an outward preferred direction, as we proposed (Fig. 1c).

Co-stratification analysis

Off SACs stratify at a particular depth in the IPL (Fig. 1b). Why this depth and not some other? From Fig. 4a, it is obvious that this depth is appropriate for wiring with BC2 and BC3a, as required by our model of DS emergence. Following this logic one step further, we wondered whether the observed dependence of contact on distance from the SAC soma might be reflected in fine aspects of SAC morphology. We hypothesized that SAC dendrites are 'tilted', moving deeper into the IPL with distance from the SAC soma. Such a change in depth would be compatible with more overlap with BC2 near the soma, and more overlap with BC3a far from the soma, as BC3a is deeper in the IPL than BC2 (Fig. 4a and Supplementary Video 1).

The hypothesized tilt turns out to exist (Fig. 5a). Very close to the SAC soma, the dendrites dive sharply into the IPL from the inner nuclear layer (INL). Surprisingly, IPL depth continues to increase as distance from the SAC soma in the tangential plane ranges from 20 to 80 μm. The slight increase is not evident in a single dendrite (Fig. 1b), but emerges from statistical averaging.

Could dendritic tilt be the cause of the observed variation in BC–SAC contact with distance (Fig. 4d)? We cannot address causality on the basis of our data, but we can test how well the tilt predicts contact variation. We computed the stratification profiles of BC types (Fig. 5a), defined as the one-dimensional density of BC surface area along the depth of the IPL. We also computed the stratification profile of SAC dendrites at various distances from the SAC soma (quartiles, Fig. 5a). Assuming that BC and SAC arborizations are statistically independent of each other, we estimated contact from 'co-stratification', defined as the integral over IPL depth of the product of BC and SAC stratification profiles (Methods).

We found that actual BC2 contact depends more strongly on distance than predicted; the slight change in IPL depth after the initial plunge appears too small to account for the large change in actual BC2 contact. In other failures of contact prediction, BC3a, BC3b and BC4 stratify at the same IPL depths (Fig. 5a), yet BC3a makes much more contact than BC3b or BC4. Also, actual BC3a contact plummets near the tips of SAC dendrites (Fig. 4d), whereas predicted contact does not change at all because the IPL depth of SAC dendrites is constant in this region (Fig. 5b). Overall, the total contact from all BC types seems low in this region (Extended Data Fig. 7d), suggesting that BCs avoid making synaptic inputs to the most distal SAC dendrites. This runs counter to the conventional belief that input synapses are uniformly distributed over the entire length of SAC dendrites[15]. The unreliability of inferring contact from co-stratification is illustrated by numerous examples of SAC dendrites that pass through BC axonal arborizations without making any contact at all (Extended Data Fig. 8).

Model of the BC–SAC circuit

We mentioned previously that BC2 lags BC3a in visual response. There is another important difference: BC3a responds more transiently to step changes in illumination, whereas BC2 exhibits more sustained responses. The implications of the sustained–transient distinction for DS can be understood using a mathematical model. The activity of a retinal neuron is often approximated as a linear spatiotemporal filtering of the visual

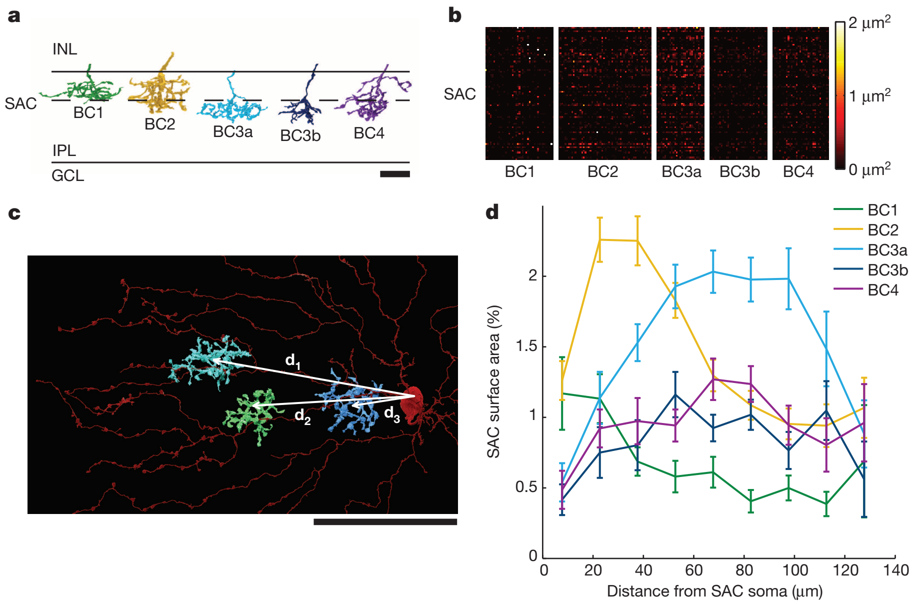


Figure 4 | BC–SAC contact. a, Off BCs were divided into five types[6,14] on the basis of IPL depth and size. Scale bar, 10 μm. **b**, Contact areas of BC–SAC pairs, sorted by BC types. **c**, Pairs were further sorted by the distance of the BC axon from the SAC soma, as measured in the tangential plane. Scale bar, 50 μm. **d**, Average BC–SAC contact versus distance, normalized to percentage of SAC surface area at that distance (Extended Data Fig. 3b). Standard error is based on the number of pairs for each BC type and distance (see Source Data for sample sizes).

stimulus followed by a nonlinearity[16,17]. Such a 'linear–nonlinear' model for the output $O(t)$ of the SAC dendrite can be written as

$$O(t)=\left[\int dxdt' W(x,t-t')I(x,t')\right]^{+} \tag{1}$$

For simplicity, the dendrite and visual stimulus $I(x,t)$ are restricted to a single spatial dimension x, and the nonlinearity is a half-wave rectification, $[z]^{+}=\max\{z,0\}$. We interpret the integral in equation (1) as the summed input from the BCs presynaptic to the SAC. The nonlinearity could arise from various biophysical mechanisms, such as synaptic transmission from SACs to other neurons. The spatiotemporal filter $W(x,t)$ is a sum of two functions,

$$W(x,t)=U_s(x)v_s(t)+U_t(x)v_t(t) \tag{2}$$

corresponding to contributions from BC2 and BC3a. The sustained temporal filter $v_s(t)$ is monophasic, whereas the transient filter $v_t(t)$ is biphasic (Fig. 6a). The spatial filter $U_s(x)$ represents the entire set of all BC2 inputs to the dendrite, and can be estimated from the BC2 contact area graph in Fig. 4d. Similarly, $U_t(x)$ can be estimated from the BC3a contact area graph. The two spatial filters are displaced relative to each other (Fig. 6a), because BC3a tends to contact SAC dendrites at more distal locations than BC2.

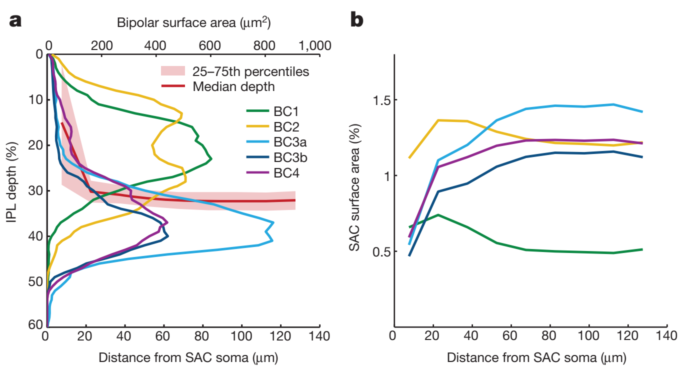


Figure 5 | BC–SAC co-stratification. a, SAC dendrites move deeper into the IPL (median depth, red line) with increasing distance from the SAC soma in the tangential plane. Stratification profiles of BC types, defined as density of surface area over the depth of the IPL. **b**, Co-stratification predictions of BC–SAC contact area versus distance from the SAC soma. The curves are normalized by SAC area at each distance, and are therefore directly comparable with those of Fig. 4d.

Each of the terms in the sum of equation (2) is said to be 'space–time separable', because it is the product of a function of space and a function of time. It was previously observed that a spatiotemporal filter $W(x,t)$ of this form can endow a model like equation (1) with DS[18,19]. This is illustrated by Fig. 6 using the fact that the convolution in equation (1) is equivalent to 'sliding' the spatiotemporal filter W in time over the stimulus I, and computing the overlap at each time. The filter $W(x,t)$ is oriented in space–time (Fig. 6a), and so also is a moving stimulus $I(x,t)$ (Fig. 6g, h). The overlap with a rightward-moving stimulus (Fig. 6h) is greater than for a leftward one (Fig. 6g), so the model exhibits DS with a rightward preferred direction.

How is DS affected by the biphasic shape of the transient temporal filter, $v_t(t)$? If we remove the negative lobe (Fig. 6c), then $v_t(t)$ will become monophasic like $v_s(t)$ and their relation closer to a simple time lag (Fig. 6d). We will refer to this model as a 'Reichardt detector', in honour of the pioneering researcher Werner Reichardt, although it more closely resembles a subunit of his model[20]. On the other hand, removing the positive lobe of $v_t(t)$ makes it monophasic but with inverted sign relative to the sustained filter (Fig. 6e). The result (Fig. 6f) resembles a DS model originally proposed by Barlow and Levick[21].

Both modified models (Fig. 6d, f) exhibit DS. In the Reichardt detector, the inputs from the two arms enhance each other for motion in the preferred direction. In the Barlow–Levick detector, the two inputs cancel each other for motion in the null direction. As our sustained–transient model (Fig. 6b) uses both mechanisms, it should exhibit more DS than either detector. Our model is related to versions of the Reichardt detector with low-pass and high-pass filters on the two arms[22].

In the original Barlow–Levick model, the negative filter corresponded to synaptic inhibition. As BCs are believed to be excitatory, negative BC input in our model represents a reduction of excitation relative to the resting level, rather than true inhibition. Signalling by reduced excitation may be possible, at least for low-contrast stimuli, as BC ribbon synapses may have a significant resting rate of transmitter release[23].

The model of equations (1) and (2) is a useful starting point for many theoretical investigations that are outside the scope of this article. For example, DS dependency on the spatial and temporal frequencies of a sinusoidal travelling wave stimulus is calculated in Supplementary Equations, and DS dependence on stimulus speed is graphed in Extended Data Fig. 9.

Discussion

In our DS model, SAC dendrites are wired to BC types with different time lags. A previous model did not distinguish between BC types, and instead relied on the time lag of signal conduction within the SAC dendrite itself[24]

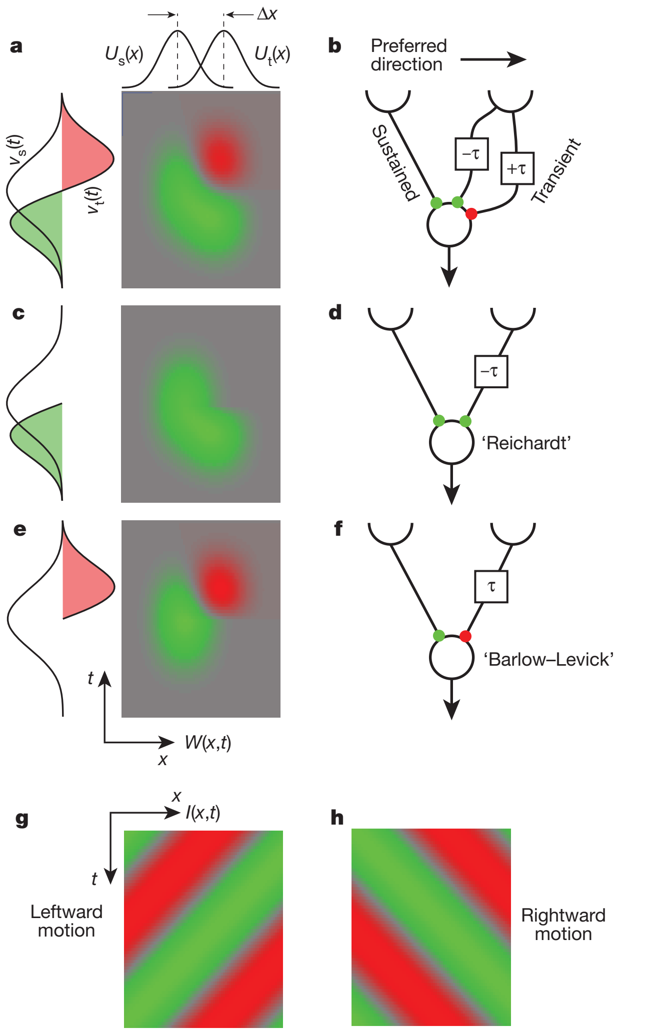


Figure 6 | Mathematical model of the BC–SAC circuit. **a**, Spatiotemporal filter of equation (2). Green is positive, red is negative, and grey is zero. **b**, The transient pathway effectively combines a positive channel that leads the sustained pathway by τ and a negative channel that lags by τ. **c**, Removing the negative channel yields a Reichardt detector (**d**). **e**, Removing the positive channel yields a Barlow–Levick detector (**f**). A moving visual stimulus $I(x,t)$ is oriented in space–time (**g**, **h**), and so are the spatiotemporal filters (**a**, **c**, **e**).

(Fig. 1d). Like most other amacrine cells, SACs lack an axon; their output synapses are found in the distal zones of their dendrites[15] (Fig. 1a, inset). Owing to dendritic conduction delay, proximal BC inputs should take longer to reach the output synapses than distal BC inputs (Fig. 1d). Therefore this time lag is also consistent with the empirical finding of an outward preferred direction. To summarize the novelty of our hypothesis, we place the time lag before BC–SAC synapses, whereas the previous model places it after BC–SAC synapses.

The postsynaptic delay model has a major weakness. If dendritic conduction were the only source of time lag, the somatic voltage would exhibit DS with an inward preferred direction, but this is inconsistent with intracellular recordings[3] (Fig. 1e). By contrast, the presynaptic delay model is compatible with approximating a SAC dendrite as isopotential (Fig. 1c), so preferred direction is predicted to be independent of the location of the voltage measurement, consistent with empirical data[3]. It may also be possible to make the postsynaptic delay model consistent with experiments by adding active dendritic conductances[4].

The presynaptic and postsynaptic delay models are not mutually exclusive. If they work together, passive cable theory suggests that presynaptic delay dominates, because estimated postsynaptic delay is much shorter than the time lag between BC2 and BC3a (Supplementary Equations). Can we gauge the relative importance of the delays empirically rather than theoretically? One way would be intracellular recording at the SAC soma of responses to visual stimulation at various dendritic locations. If postsynaptic delay dominates, then response latency will grow with distance of the visual stimulus from the soma. If presynaptic delay dominates, then distal stimulation will evoke somatic responses with shorter latency than proximal stimulation. This prediction may seem counterintuitive, but is an obvious outcome of our model.

Many other models of DS emergence in SACs invoke inhibition as well as excitation[25–28]. We have focused on excitatory mechanisms, as blocking inhibition does not abolish DS[3]. However, inhibition may have the effect of enhancing DS, and its role should be investigated further.

This work focused on Off BC–SAC circuitry. An analogous sustained–transient distinction can also be made for On BC types[7,8]. It remains to be seen whether their connectivity with On SACs depends on distance from the soma. If this turns out to be the case, then the model of Fig. 6 could serve as a general theory of motion detection by both On and Off SACs. The model filter of Fig. 6a also resembles the spatiotemporal receptive field of the J type of ganglion cell (see Fig. 3b of ref. 29).

Neural activity imaging[30] and connectomic analysis[31] have recently identified a plausible candidate for the site of DS emergence in the fly visual system. If our theory is correct, then the analogies between insect and mammalian motion detection[1] are more far-reaching than previously suspected, with fly T4 and T5 cells corresponding to On and Off SAC dendrites in both connectivity and function.

A glimmer of space–time wiring specificity can even be seen in the structure of the SAC itself. As BC types with different time lags arborize at different IPL depths, IPL depth can be regarded as a time axis. Therefore, the slight tilt of the SAC dendrites in the IPL (Fig. 5a) could be related to the orientation of the SAC receptive field in space–time (Fig. 6a). However, dendritic tilt alone is not sufficient to predict our model, as co-stratification sometimes fails to predict contact (Figs 4d and 5b). For example, co-stratification predicts strong BC4 connectivity to distal SAC dendrites. This would favour an inward preferred direction, contrary to what is observed, because BC2 leads (not lags) BC4 in visual responses[7].

The idea that contact (or connectivity) can be inferred from co-stratification is sometimes known as Peters' rule[32], and has also been applied to estimate neocortical connectivity[33–35]. The present work shows that fairly subtle violations of Peters' rule may be important for visual function. Previous research suggests that On–Off direction-selective ganglion cells inherit their DS from SAC inputs owing to a strong violation of Peters' rule[9,36–38].

Our findings were made possible by using artificial intelligence to reduce the amount of human effort required for 3D reconstruction of neurons. Even after the labour savings, our research required great human effort from a handful of paid workers in the laboratory and a large number of volunteers through EyeWire. Our experiences do not support claims that the 'wisdom of the crowd' should replace experts[39]. Instead, EyeWire depends on cooperation between laboratory experts and online amateurs (Methods). Furthermore, some amateurs developed remarkable expertise and were promoted to increasingly sophisticated roles within the EyeWire community (Supplementary Notes). We believe that crowd wisdom requires amplifying the expert voices within the crowd, and also empowering individuals to become experts. Fortunately, such goals are well-matched to the game format.

The EyeWire artificial intelligence was based on a deep convolutional network[40,41]. Similar networks have been successfully applied to serial electron microscopy images obtained using conventional staining techniques that mark intracellular organelles[42]. Extending EyeWire to such images, in which synapses are clearly visible, would enable a true connection analysis that goes beyond the contact and co-stratification analyses used here.

Our work demonstrates that reconstructing a neural circuit can provide surprising insights into its function. Much more will be learned as reconstruction speed grows. The combination of crowd and artificial intelligence promises a continuous upward path of improvement, as human input from the crowd is not only useful for generating neuroscience

discoveries, but also for making the artificial intelligence more capable through machine learning.

Note added in proof: Further evidence that BC axons exhibit little or no DS appeared while this paper was in press[43].

METHODS SUMMARY

A convolutional network was trained to detect neural boundaries via the MALIS procedure[40] and CNPKG (https://github.com/srinituraga/cnpkg/), which is based on Cortical Network Simulator[44]. The convolutional network was applied to the e2198 data set, which was then segmented into supervoxels by a modified version of the watershed algorithm. Paid workers and volunteer EyeWirers reconstructed neurons in 3D by assembling supervoxels. The retina was computationally flattened, reconstructed neurons were classified by their structural properties, and contact and co-stratification were analysed by custom Matlab and C++ code.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 13 October 2013; accepted 10 March 2014.

Published online 4 May 2014.

1. Borst, A. & Euler, T. Seeing things in motion: models, circuits, and mechanisms. *Neuron* **71,** 974–994 (2011).
2. Vaney, D. I., Sivyer, B. & Taylor, W. R. Direction selectivity in the retina: symmetry and asymmetry in structure and function. *Nature Rev. Neurosci.* **13,** 194–208 (2012).
3. Euler, T., Detwiler, P. B. & Denk, W. Directionally selective calcium signals in dendrites of starburst amacrine cells. *Nature* **418,** 845–852 (2002).
4. Hausselt, S. E., Euler, T., Detwiler, P. B. & Denk, W. A dendrite-autonomous mechanism for direction selectivity in retinal starburst amacrine cells. *PLoS Biol.* **5,** e185 (2007).
5. Yonehara, K. *et al.* The first stage of cardinal direction selectivity is localized to the dendrites of retinal ganglion cells. *Neuron* **79,** 1078–1085 (2013).
6. Wässle, H., Puller, C., Müller, F. & Haverkamp, S. Cone contacts, mosaics, and territories of bipolar cells in the mouse retina. *J. Neurosci.* **29,** 106–117 (2009).
7. Baden, T., Berens, P., Bethge, M. & Euler, T. Spikes in mammalian bipolar cells support temporal layering of the inner retina. *Curr. Biol.* **23,** 48–52 (2013).
8. Borghuis, B. G., Marvin, J. S., Looger, L. L. & Demb, J. B. Two-photon imaging of nonlinear glutamate release dynamics at bipolar cell synapses in the mouse retina. *J. Neurosci.* **33,** 10972–10985 (2013).
9. Briggman, K. L., Helmstaedter, M. & Denk, W. Wiring specificity in the direction-selectivity circuit of the retina. *Nature* **471,** 183–188 (2011).
10. Lintott, C. J. *et al.* Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey. *Mon. Not. R. Astron. Soc.* **389,** 1179–1189 (2008).
11. Cooper, S. *et al.* Predicting protein structures with a multiplayer online game. *Nature* **466,** 756–760 (2010).
12. Fiala, J. C. Reconstruct: a free editor for serial section microscopy. *J. Microsc.* **218,** 52–61 (2005).
13. Von Ahn, L. & Dabbish, L. Labeling images with a computer game. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* 319–326 (ACM, 2004).
14. Helmstaedter, M. *et al.* Connectomic reconstruction of the inner plexiform layer in the mouse retina. *Nature* **500,** 168–174 (2013).
15. Famiglietti, E. V. Synaptic organization of starburst amacrine cells in rabbit retina: analysis of serial thin sections by electron microscopy and graphic reconstruction. *J. Comp. Neurol.* **309,** 40–70 (1991).
16. Berry, M. J., II & Meister, M. Refractoriness and neural precision. *J. Neurosci.* **18,** 2200–2211 (1998).
17. Baccus, S. A. & Meister, M. Fast and slow contrast adaptation in retinal circuitry. *Neuron* **36,** 909–919 (2002).
18. Watson, A. B. & Ahumada, A. J. Jr. Model of human visual-motion sensing. *J. Opt. Soc. Am. A* **2,** 322–341 (1985).
19. Adelson, E. H. & Bergen, J. R. Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am. A* **2,** 284–299 (1985).
20. Reichardt, W. in *Sensory Communication* (ed. Rosenblith, W. A.) 303–317 (MIT Press, 1961).
21. Barlow, H. B. & Levick, W. R. The mechanism of directionally selective units in rabbit's retina. *J. Physiol. (Lond.)* **178,** 477–504 (1965).
22. Borst, A., Reisenman, C. & Haag, J. Adaptation of response transients in fly motion vision. II: model studies. *Vision Res.* **43,** 1311–1324 (2003).
23. Lagnado, L., Gomis, A. & Job, C. Continuous vesicle cycling in the synaptic terminal of retinal bipolar cells. *Neuron* **17,** 957–967 (1996).
24. Tukker, J. J., Taylor, W. R. & Smith, R. G. Direction selectivity in a model of the starburst amacrine cell. *Vis. Neurosci.* **21,** 611–625 (2004).
25. Borg-Graham, L. J. & Grzywacz, N. M. in *Single Neuron Computation* (eds McKenna, T., Davis, J. & Zornetzer, S. F.) Ch. 13, 347–76 (Academic San Diego, 1992).
26. Gavrikov, K. E., Dmitriev, A. V., Keyser, K. T. & Mangel, S. C. Cation–chloride cotransporters mediate neural computation in the retina. *Proc. Natl Acad. Sci. USA* **100,** 16047–16052 (2003).
27. Münch, T. A. & Werblin, F. S. Symmetric interactions within a homogeneous starburst cell network can lead to robust asymmetries in dendrites of starburst amacrine cells. *J. Neurophysiol.* **96,** 471–477 (2006).
28. Lee, S. & Zhou, Z. J. The synaptic mechanism of direction selectivity in distal processes of starburst amacrine cells. *Neuron* **51,** 787–799 (2006).
29. Kim, I.-J., Zhang, Y., Yamagata, M., Meister, M. & Sanes, J. R. Molecular identification of a retinal cell type that responds to upward motion. *Nature* **452,** 478–482 (2008).
30. Maisak, M. S. *et al.* A directional tuning map of *Drosophila* elementary motion detectors. *Nature* **500,** 212–216 (2013).
31. Takemura, S. Y. *et al.* A visual motion detection circuit suggested by *Drosophila* connectomics. *Nature* **500,** 175–181 (2013).
32. Braitenberg, V. & Schüz, A. *Cortex: Statistics and Geometry of Neuronal Connectivity* 2nd edn (Springer Berlin, 1998).
33. Kalisman, N., Silberberg, G. & Markram, H. Deriving physical connectivity from neuronal morphology. *Biol. Cybern.* **88,** 210–218 (2003).
34. Binzegger, T., Douglas, R. J. & Martin, K. A. C. A quantitative map of the circuit of cat primary visual cortex. *J. Neurosci.* **24,** 8441–8453 (2004).
35. Stepanyants, A. & Chklovskii, D. B. Neurogeometry and potential synaptic connectivity. *Trends Neurosci.* **28,** 387–394 (2005).
36. Fried, S. I., Münch, T. A. & Werblin, F. S. Mechanisms and circuitry underlying directional selectivity in the retina. *Nature* **420,** 411–414 (2002).
37. Yonehara, K. *et al.* Spatially asymmetric reorganization of inhibition establishes a motion-sensitive circuit. *Nature* **469,** 407–410 (2010).
38. Wei, W., Hamby, A. M., Zhou, K. & Feller, M. B. Development of asymmetric inhibition underlying direction selectivity in the retina. *Nature* **469,** 402–406 (2011).
39. Surowiecki, J. *The Wisdom of Crowds* (Anchor, 2005).
40. Turaga, S., Briggman, K., Helmstaedter, M., Denk, W. & Seung, H. S. in *Advances in Neural Information Processing Systems 22*, 1865–1873 (2009).
41. Turaga, S. C. *et al.* Convolutional networks can learn to generate affinity graphs for image segmentation. *Neural Comput.* **22,** 511–538 (2010).
42. Ciresan, D. *et al.* in *Advances in Neural Information Processing Systems 25*, 2852–2860 (2012).
43. Park, S. J. H., Kim, I. -J., Looger, L. L., Demb, J. B. & Borghuis, B. G. Excitatory synaptic inputs to mouse on-off direction-selective retinal ganglion cells lack direction tuning. *J. Neurosci.* **34,** 3976–3981 (2014).
44. Mutch, J., Knoblich, U. & Poggio, T. *CNS: a GPU-Based Framework for Simulating Cortically-Organized Networks* Tech. Rep. MIT-CSAIL-TR-2010-013/CBCL-286 (MIT, 2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements This research was made possible by funding from the Gatsby Charitable Foundation, the Howard Hughes Medical Institute, the Human Frontier Science Program, an anonymous donor, and the National Institutes of Health. K.L. was supported by a Samsung Scholarship. Support from the AWS Research Grants Program gave EyeWire global reach through Amazon Cloudfront. We thank K. Briggman for providing the e2198 data set. J. Mutch created the CNS framework on which CNPKG is based. D. Jia, R. Shearer, and B. Warne assisted in early stages of software development, and W. Silversmith with recent modifications. R. Prentki, L. Trawinski, M. Sorek, A. Ostojic, C. David, R. Avery, S. Temple, A. Bost, M. Greenstein and M. Evans worked in the laboratory to reconstruct neurons, and the first six also served as GrimReaper and hosted EyeWire competitions. Additional reconstructions were provided by R. Han, M. Gavrin, G. Lu, A. Ortiz and D. Udvary. All were trained by R. Prentki, who also created training videos for EyeWirers. We are grateful to A. Norton for 3D renderings, and to E. Almeida for EyeWire graphics. We acknowledge discussions with T. Baden, M. Berry, B. Borghuis, A. Borst, E. J. Chichilnisky, D. Chklovskii, D. Clark, J. Demb, T. Euler, M. Helmstaedter, A. Huberman, S. Lee, R. Masland, J. Sanes and Z. Zhou.

Author Contributions J.S.K. created algorithms, software and procedures for crowd intelligence and learning, and applied them to generate neuron reconstructions. J.S.K. and M.J.G. classified bipolar cells. M.J.G. analysed contact and co-stratification, aided by code from A.Z. and input from W.D. H.S.S. devised the model with help from B.F.B. and M.C. S.C.T. trained the convolutional network. M.P. and M.B. implemented software and algorithms created by A.Z. for interactive segmentation and 3D visualization, with guidance from S.C.T. M.R. created the EyeWire game and M.B. its data infrastructure. K.L. quantified EyeWirer accuracy and learning. A.R. mobilized and studied the EyeWire community. EyeWirers reconstructed neurons and built extensions to EyeWire. H.S.S. wrote the paper with help from J.S.K., M.J.G. and A.R.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.S.S. (sseung@princeton.edu).

METHODS

We worked with the e2198 data set[9] rather than the e2006 data set[14] because e2198 is large enough to encompass entire SAC dendrites (~150 μm). All dimensions are uncorrected for tissue shrinkage, which was previously estimated at 14% by comparison of two-photon and serial electron microscopy images[14].

Machine learning. The boundaries between neurons in subvolumes of the e2198 and e2006 data sets were manually traced. Using this as ground truth, a convolutional network was trained to detect boundaries between neurons using the MALIS method[40]. The convolutional network had the same architecture as one used previously[14], and produced as output an affinity graph connecting nearest neighbour voxels[41]. Any subvolume of e2198 could be oversegmented by applying a modified watershed algorithm to the appropriate subgraph. The regions of the oversegmentation are called supervoxels.

Reconstruction by workers. A team of part-time workers, numbering about half a dozen at any given time, reconstructed neurons using a more sophisticated version of the EyeWire interface. Workers were hired on the basis of an interview and a test of software use passed by three-quarters of the applicants. They were trained for 40–50 h before generating reconstructions used for research. Their skills typically improved for months or even years after the initial training period, and were superior to those of professional neuroscientists without reconstruction experience.

As with EyeWire, the task of reconstructing an entire neuron was divided into subtasks, each of which involved reconstructing the neuron within a subvolume starting from a supervoxel 'seed'. However, the subvolumes were roughly 100 times larger than EyeWire cubes, and only two workers were assigned to each subvolume.

In the first stage of error correction, disagreements were detected by computer, and resolved by one of the two workers, or a third worker. The third occasionally detected and corrected errors that were not disagreements between the first two. Most disagreements were the result of careless errors, and were easily resolved. More rarely, there were disagreements caused by fundamental ambiguities in the image. These locations were noted for later examination in a further stage of error correction.

This second stage relied on 3D reconstructions of entire neurons assembled from multiple subvolumes and inspected by one of the authors (J.S.K.). Suspicious branches or terminations, as well as overlaps between reconstructions of different neurons were detected. The original image was re-examined at these locations to check for errors. The process was repeated until no further errors could be detected.

The precision of our final reconstruction relative to the truth is probably comparable to the precision of the penultimate reconstruction relative to the final reconstruction, 0.99 for SACs and 0.96 for BCs. Recall is probably somewhat poorer, because missing branches are more difficult to detect than superfluous branches. Recall must be reasonably good for SACs, as missing branches would be detected by deviations from the typical SAC shape and radius.

Reconstruction by EyeWirers. Some reconstruction errors slip past the consensus mechanism. These are detected through visual inspection of an 'overview' mode, which displays 3D renderings of entire neurons currently under reconstruction (Extended Data Fig. 1b). False branches become obvious once they are long enough, and are reported by EyeWirers through chat or email. They are chopped off by GrimReaper, a special EyeWirer played by laboratory experts endowed with the superpower of overruling the consensus. GrimReaper also extends branches that have terminated prematurely. Correction by GrimReaper is similar to the second stage of error correction described above, so the final reconstruction presumably has similar accuracy.

SAC reconstructions are extremely difficult for two reasons: (1) SAC dendrites are very thin and may falsely appear to terminate, owing to limited spatial resolution and imperfect staining; and (2) the interiors of many SAC boutons contained irregular darkenings, which could falsely appear like cellular boundaries. (The reason for the darkening is unclear, as the extracellular staining procedure was not intended to mark intracellular structures.)

Novices tend to prematurely terminate SAC dendrites. Experts know that most cubes do not contain termination points, and therefore try harder to find continuations, using a variety of sophisticated search strategies. GrimReaper is also allowed to view how the cube fits into the entire reconstructed neuron. This additional spatial context can be used to disambiguate difficult cubes, given knowledge of the typical appearance of a SAC.

Before learning in normal gameplay (Fig. 2d), all EyeWirers are required to go through a training session immediately after registering for the site. This consists of a sequence of tutorial cubes, each of which was previously coloured by an expert (Extended Data Fig. 1c). Each cube teaches through instructions and per-click feedback about accuracy based on comparing the EyeWirer's selections with those of the expert. After submitting a tutorial cube, the EyeWirer is given a chance to view mistakes.

Accuracy is monitored on a weekly basis by computing the precision and recall of each EyeWirer with respect to the truth, defined as neuron reconstructions based on EyeWire consensus followed by GrimReaper corrections. Less accurate EyeWirers are given less weight in the vote.

Players' daily, weekly and monthly scores are publicly displayed on a leaderboard (Extended Data Fig. 1b, right), motivating players to excel through competition. Players communicate with each other through online 'chat' (Extended Data Fig. 1b, left) and discussion forums.

A 'beta test' version of EyeWire was deployed in February 2012 and attracted a small group of users, who helped guide software development. EyeWire officially launched in December 2012.

Reconstruction of Off SACs. Off SACs were recognized by their somata in the INL, narrow IPL stratification at roughly one third of the depth from the INL to the ganglion cell layer, and characteristic 'starburst' appearance (Fig. 1a).

Off SACs were reconstructed by: (1) forward tracing from the soma to dendritic tips; and (2) backward tracing from varicosities on candidate SAC dendrites to the soma. In the forward method, a candidate SAC soma was identified as a supervoxel with a characteristic pattern of dendritic stubs bearing spiny protrusions. By the time reconstruction progressed to approximately half of the average SAC radius, an Off SAC could be conclusively recognized by its starburst shape and narrow stratification at the appropriate IPL depth. More than 90% of candidates turned out to be SACs.

In the backward method, we located a thin dendrite with varicosities at the appropriate IPL depth. This was reconstructed back to the soma, and then the rest of the dendrites were reconstructed from the soma to the tips. The cell could be discarded at any point during this process, if its dendrites escaped from the appropriate IPL depth or failed to exhibit the proper morphological characteristics. Less than 25% of initial candidates ended up confirmed as SACs.

In total, 79 Off SACs were reconstructed, 39 by forward tracing and 52 by backward tracing. This is more than half the entire population in e2198, judging from the published density[45]. After candidates were identified by one of the authors (J.S.K.), reconstructions were performed by laboratory workers (59 cells) or by EyeWirers (29 cells). Both pairs of numbers sum to more than 79, because the sets overlapped (12 for forward/backward, 9 for workers/EyeWirers).

In March 2012, laboratory workers began reconstruction of SACs. In March 2013, EyeWirers were invited to the 'Starburst Challenge', a sequence of tutorial cubes drawn from SACs. Those who passed with sufficient accuracy were an elite group allowed to reconstruct SACs (Supplementary Information). EyeWirers eventually shouldered most of the burden of SAC reconstruction, with only 8% of SAC cubes needing correction by GrimReaper. This enabled laboratory workers to shift their focus to BCs, as described below.

Reconstruction of Off BCs. The somata of Off BCs were generally outside e2198, which extended only partially into the INL (Fig. 1 of ref. 9). The trunks of candidate BC axons were located in the interstices of the INL, and followed into the IPL. If the axons arborized in the Off region of the INL, they were fully reconstructed. Cells that violated known BC structures were identified as amacrine cells and discarded[14].

BC axons were difficult to reconstruct owing to poor staining, and their highly irregular shapes. They could not be accurately reconstructed (either by online volunteers or laboratory experts) within the $256 \times 256 \times 256$ cubes of EyeWire, which were too small to provide sufficient spatial context. Therefore BCs were reconstructed only by laboratory workers using the large subvolumes mentioned above.

Coordinate system. For more precise quantification of structural properties, a new coordinate system was defined by applying a nonlinear transformation to neurons so as to flatten the IPL and make it perpendicular to one of the coordinate axes. The nonlinear transformation was found by the following steps. First a global planar approximation to the Off SAC surface was computed. Then the centroid of all the SACs was projected onto this global plane to define the origin of the coordinate system. The projection was along the coordinate axis of the e2198 volume closest in direction to the light axis.

To correct for curvature, an azimuthal equidistant projection[46] of the Off SAC surface onto the global plane was made about the origin. Then local planar approximations to the SAC surface were computed in the neighbourhoods of every node in a triangular lattice. At each point in a triangle, the SAC surface was approximated by computing the mean of the planar approximations (as quaternions with yaw constrained to be zero) for the triangle's vertices, weighted by distance of the point from the vertices.

The Off SACs were defined as 32% IPL depth. We also reconstructed a few On SACs, and defined them as 62%. These choices placed the edge of the INL at 0%. Structural properties of all cells were computed on the basis of locations of their surface voxels after transformation into the new coordinates.

Classification of Off bipolar cells. BC stratification profiles were computed by dividing surface voxels into 100 bins spanning 0 to 100% IPL depth. Classification

into cell types was done by using methods similar to those described previously[14]. The BCs were split into shallow (BC1/2) and deep (BC3/4) clusters using the 75th percentile depth of the stratification profile. The BC1/2 cluster was further subdivided into two clusters by stratification width, defined as the difference between 75th and 25th percentile depths. On the basis of cells per square millimetre (Extended Data Fig. 6f), we inferred that the wider cluster was BC2 and the narrower cluster was BC1. These two types were originally defined by molecular criteria[6], and our inferred correspondence with structural definitions is transposed relative to a previous report[14]. The BC3/4 cluster was subdivided into BC4 and BC3 by the 10th percentile depth, because the molecularly defined BC4 stratifies closer to the INL[6]. Finally, BC3 was subdivided into BC3a and BC3b on the basis of axonal arborization volume, with BC3a having the larger axonal volume. Each of the above subdivision steps was based on a feature with a roughly bimodal histogram (Extended Data Fig. 5).

The result still contained a small number of classification errors, detected when adjacent BCs of the same type overlapped enough to violate the mosaic property. Corrections were made by an automatic algorithm that greedily swapped cells from one cluster to another such that the total overlap between convex hulls of cells of a given type was minimized. Two swaps were vetoed by an expert (J.S.K.) on the basis of morphological features. In all, six cells were swapped within BC1/2 and 13 within BC3/4. In the final classification, 41, 56, 29, 35 and 34 BCs were identified as types 1, 2, 3a, 3b and 4, respectively (Extended Data Fig. 6). Cells that violated the mosaic of all types (7) or had irregular stratification profiles (9) were discarded as possible reconstruction errors or amacrine cells.

Contact analysis. Edges of the affinity graph connecting BC with SAC voxels were defined as BC–SAC contact edges. For each pair, the sum of the edges yielded an estimate of contact area. The Euclidean distance separating each BC–SAC pair was computed after projecting their centres onto the SAC plane. Centres of SAC somata were manually annotated, and centres of BC arborizations were computed as the centroids of their surface voxels. The pairs were binned by distance of the BC from the SAC soma. For every pair in a bin, the fraction of SAC surface area devoted to BC–SAC contact within the convex hull of the BC was computed as the ratio of BC–SAC contact edges to SAC surface edges within the convex hull. The latter was estimated by the number of SAC surface voxels multiplied by a geometric conversion factor of 1.4 SAC surface edges per surface voxel. (This factor was estimated by dividing the total number of SAC surface edges by the total number of SAC surface voxels in the volume.) BC–SAC pairs with fewer than 10,000 SAC surface voxels inside the hull were excluded from the computation to reduce the effect of fluctuations. The ratios for BCs of the same type were averaged for each distance bin and multiplied by a mosaic overlap factor to yield the values in Fig. 4d. The mosaic overlap factor represents the extent to which neighbouring convex hulls overlap one another, which varies by cell type. This factor was computed by dividing the sum of the hull areas for each cell by the area of the union of hulls for each cell type. For absolute rather than fractional areas, edges in the affinity graph were converted to area in μm^2, using the conversion factor of 291.5 μm^2 per edge. This factor averages over the different edge orientations and compensates for voxelization effects. A result very similar to Fig. 4d can also be obtained by an alternative method that is simpler but does not yield error bars (Extended Data Fig. 7c).

Co-stratification analysis. All SAC surface voxels were binned by distance from the soma centre in the SAC plane. Within each bin, the stratification profile was computed as for the BCs. The quartiles (median and 25th and 75th percentiles) are graphed in Fig. 5a. The prediction of contact from co-stratification is based on the following formalism.

We define the arborization density $\rho_a(\mathbf{r})$ as the surface area per unit volume at location $\mathbf{r}$ of a type a cell with soma centred at the origin. Its integral $\int dxdydz\rho_a(\mathbf{r})$ is the total surface area of the arborization. We assume that the contact density received by one cell of type a from all cells of type b is equal to

$$c_{ab}(\mathbf{r}) = \rho_a(\mathbf{r}) \sum_i \rho_b(\mathbf{r} - \mathbf{r}_{bi}) \tag{3}$$

The sum over the b mosaic can be approximated by a function that is independent of x and y,

$$\sum_i \rho_b(\mathbf{r} - \mathbf{r}_{bi}) \approx \sigma_b s_b(z) \tag{4}$$

where σ_b is the number of type b neurons per retinal area and

$$s_b(z) = \int dxdy\rho_b(x,y,z) \tag{5}$$

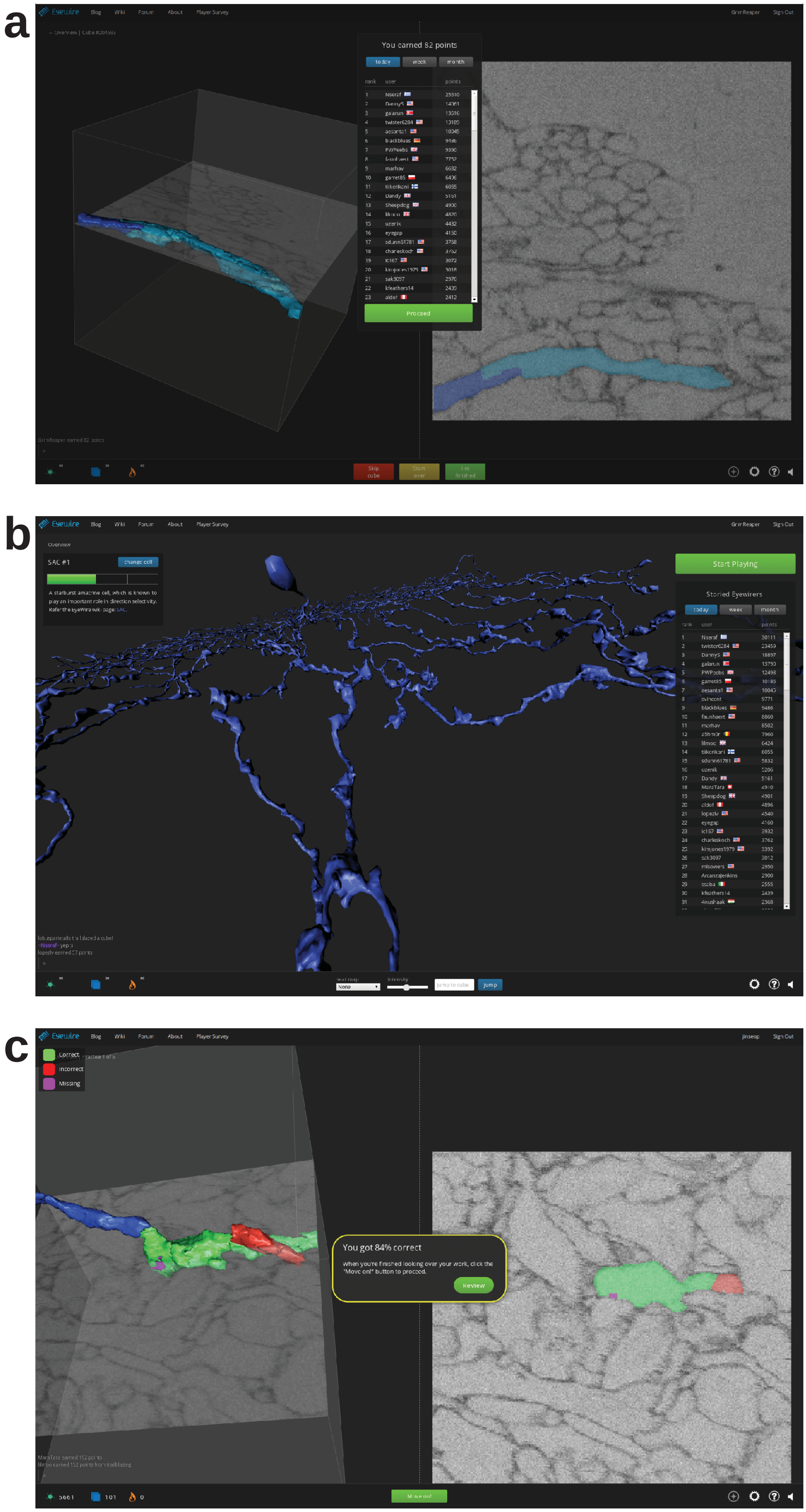
is the stratification profile of a cell of type b. The SAC arborization density is assumed radially symmetric,

$$\rho_{\rm SAC}(\mathbf{r}) = \rho_{\rm SAC}(\sqrt{x^2+y^2}, z),$$

where $\rho_{\rm SAC}(\mathbf{r})$ can be regarded (up to normalization) as the SAC stratification profile as a function of distance $r = \sqrt{x^2+y^2}$ from the SAC soma. Integrating the contact density (3) and normalizing yields the fraction $\phi_b(r)$ of SAC area contacted by cell type b as a function of r,

$$\phi_b(r) = \sigma_b \frac{\int dz\rho_{\rm SAC}(r,z)s_b(z)}{\int dz\rho_{\rm SAC}(r,z)} \tag{6}$$

45. Jeon, C.-J., Strettoi, E. & Masland, R. H. The major cell populations of the mouse retina. *J. Neurosci.* **18,** 8936–8946 (1998).
46. Snyder, J. P. *Map Projections–A working Manual* 1395 (USGPO, 1987).

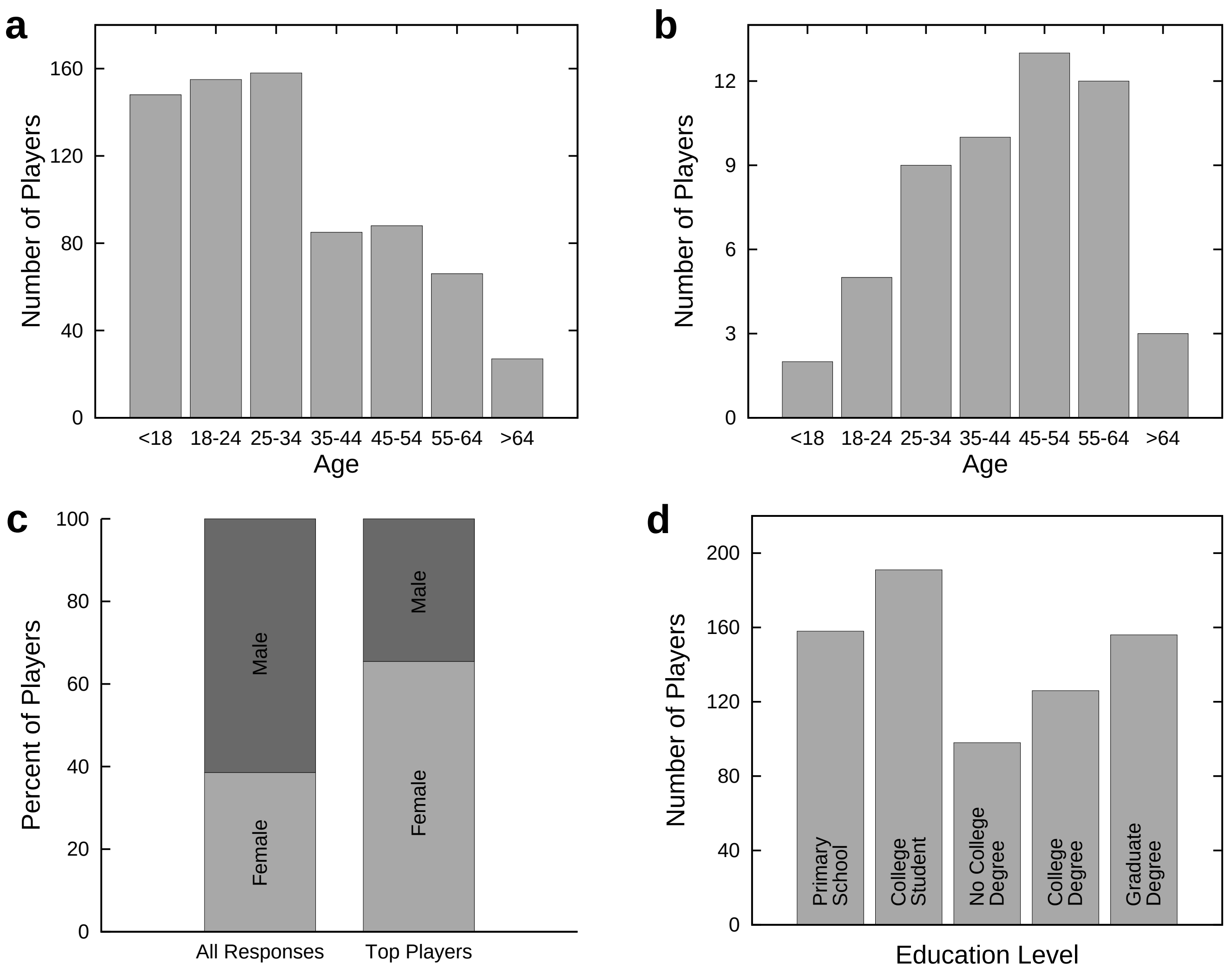


Extended Data Figure 1 | EyeWire screenshots. **a**, Numerical score after gameplay of a cube, with leaderboard below. **b**, Overview mode with neuron under reconstruction (centre), global chat (bottom left), progress bar for neuron (top left), leaderboard (right), settings and help (bottom right). **c**, Tutorial play.

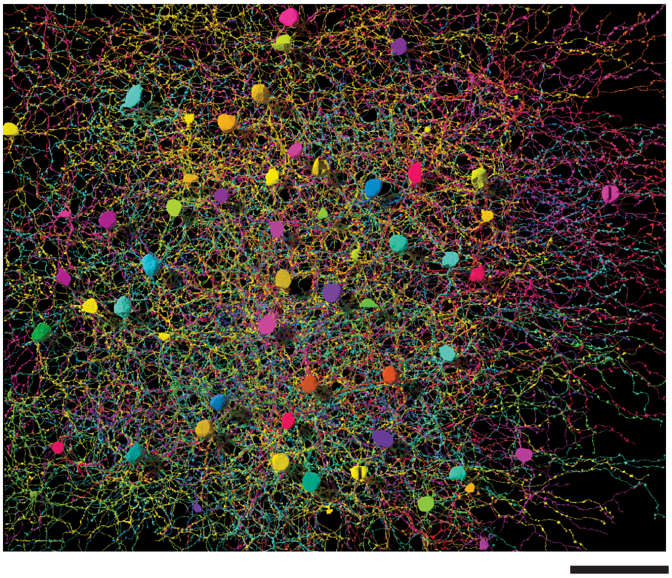
Username* (free text)
Gender*
    Male/Female
Age* (free text)
Location*
    City, State/Province
    Country
Are you…
    White or Caucasian
    Asian
    African American or Black
    American Indian or Alaska Native
    Hispanic
    Pacific Islander
Education*
    Middle School
    High School - current student
    High School
    Some College - current student
    Some College - not currently a student
    Finished College (Undergrad)
    Some Graduate School - current Masters student
    Masters -- Finished Degree
    Some Graduate School - current PhD student
    PhD -- Finished Degree
    MD/DO
Occupation* (free text)
Do you have prior experience in neuroscience?*^
    Yes/No
    If yes, please explain.*^
How long do you play EyeWire each week?*
    Less than 1 hour/More than 1 hour
If you play for more than 1 hour per week, how long do you play?
    1 to 2 hours
    3 to 5 hours
    6 to 10 hours
    11 to 20 hours
    21 to 30 hours
    31 to 40 hours
    41 to 50 hours
    More than 50 hours
What scientific purpose does EyeWire serve? (free text)
Why do you play EyeWire? (free text)
How did you discover EyeWire? (free text)
If you could add one feature to EyeWire, what would it be? (free text)
Anything else you would like to add? (free text)

Survey launch date: April 14, 2013. *required question, ^question added on 7/7/2013

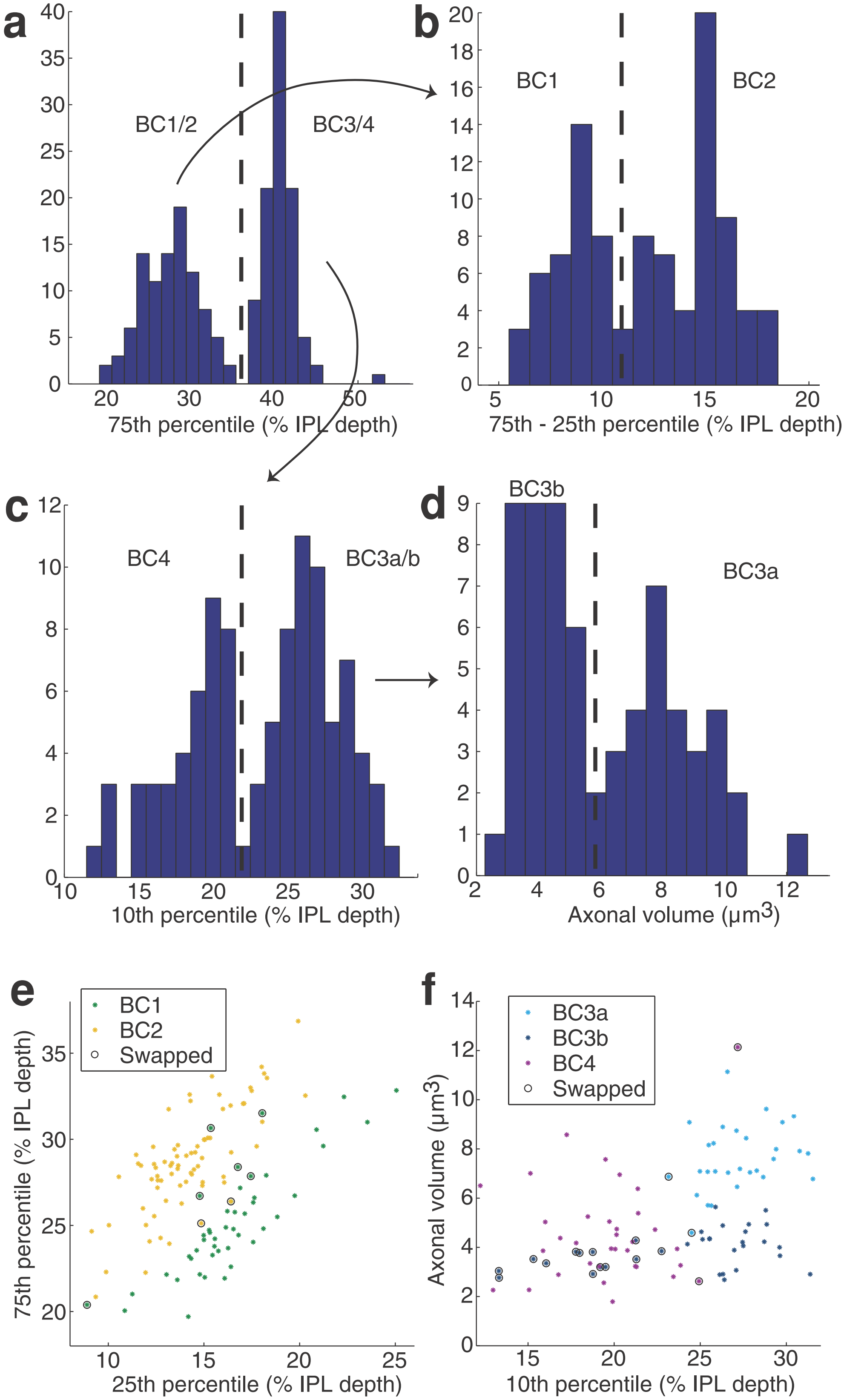
Extended Data Figure 2 | Questionnaire administered to EyeWirers.



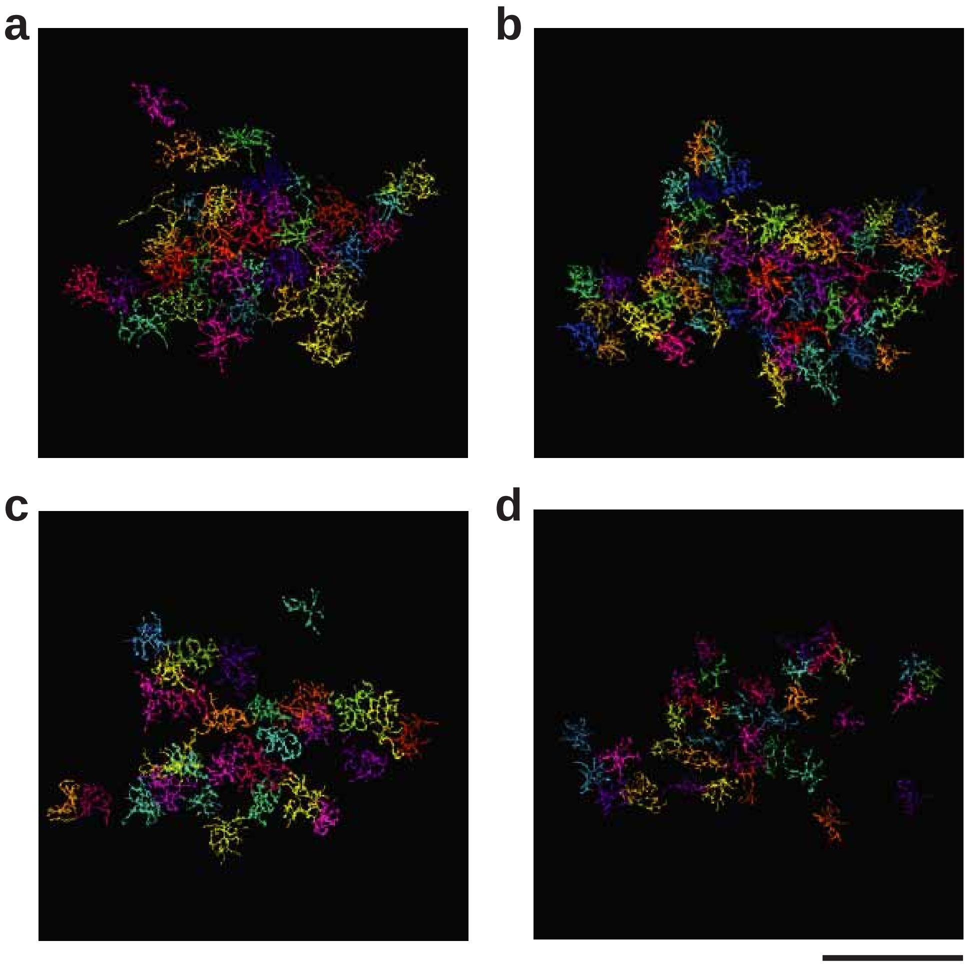
Extended Data Figure 3 | EyeWire demographics. **a**, **b**, Data based on 729 responses to the questionnaire in Extended Data Fig. 2. Age distribution of (**a**) all respondents and (**b**) those among the top 100 players ranked by number of cubes submitted. **c**, Gender distribution of all respondents and those among the top 100 players. **d**, Distribution of educational levels.

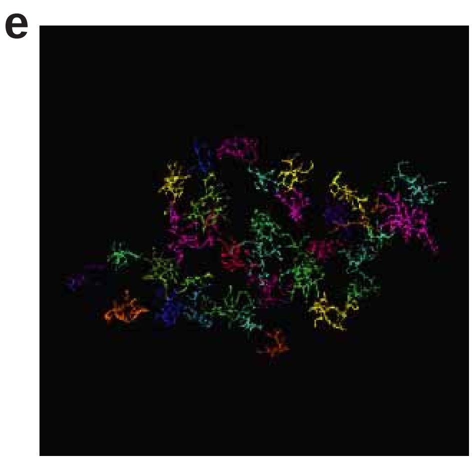


Extended Data Figure 4 | Entirety of reconstructed SACs. Only the central region of this plexus of SAC dendrites is portrayed in Fig. 3b. Scale bar, 50 μm.



Extended Data Figure 5 | Clustering procedure for BCs. **a**, Cells were divided by the 75th percentile of their stratification profiles. **b**, The shallow cluster BC1/2 was separated into BC1 and BC2 using stratification width, defined as the difference between 75th and 25th percentiles. **c**, The deep cluster BC3/4 was divided by 10th percentile into BC4 and BC3. **d**, BC3 was divided by axonal volume to yield BC3a and BC3b. Scatter plots of the BC1/2 (**e**) and BC3/4 (**f**) divisions show swaps made to eliminate mosaic violations. No swaps between BC1/2 and BC3/4 were needed.

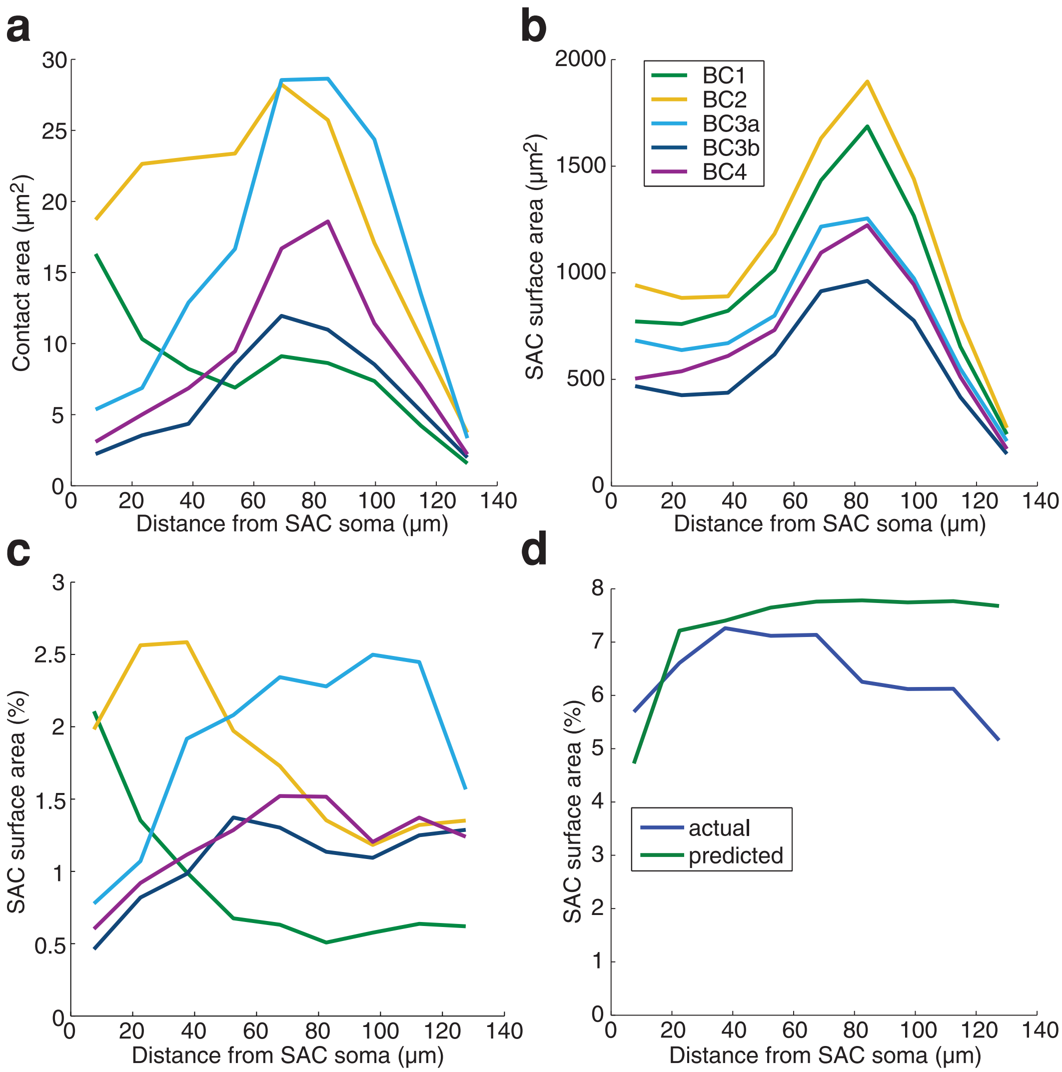




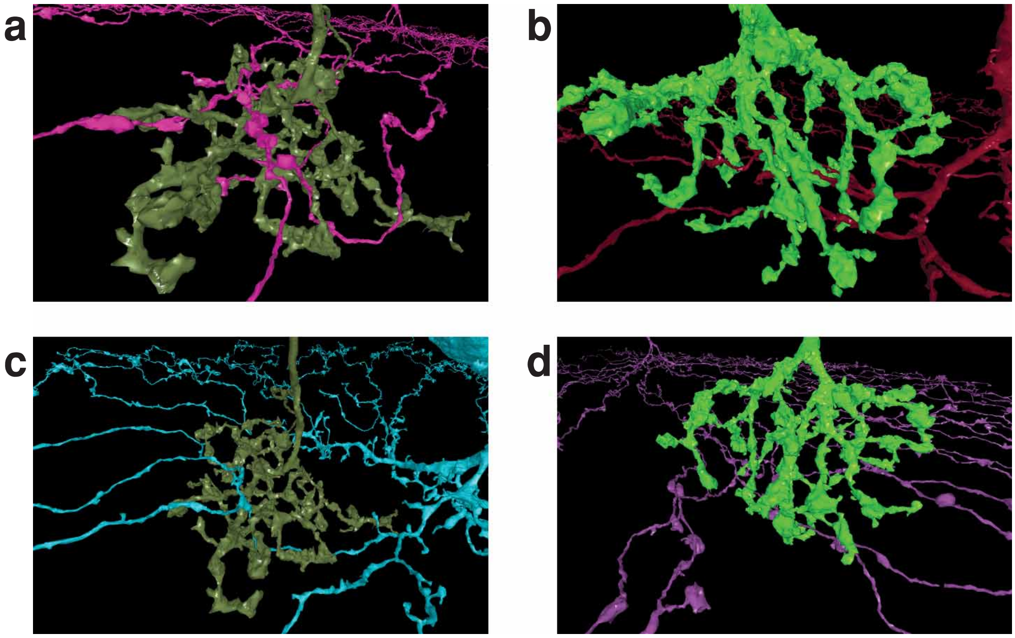
f

Type	n	Hull Area	Density	Overlap
BC1	41	382±104	3652	40%
BC2	56	297±79	4485	33%
BC3a	29	363±103	3236	18%
BC3b	35	206±57	5418	11%
BC4	34	266±114	4303	15%

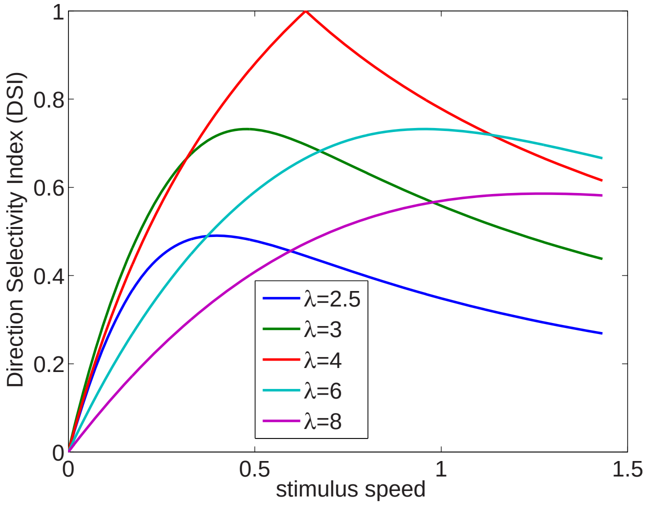
Extended Data Figure 6 | Mosaics of Off BC types. a–e, Reconstructed BCs of types 1, 2, 3a, 3b and 4 (**a** through **e**, respectively). BC1/2 mosaics appear complete. BC3/4 mosaics show some gaps, probably because some thin axons were missed in the INL (Methods). Scale bar, 50 μm. **f**, Statistics of BC types. Means and standard deviation of the hull area (area of the convex hull around the cell) are in μm^2. Type densities are the number of cells (n) divided by the area of the union of hulls of that cell type, and are in cells per mm^2 without compensation for tissue shrinkage (Methods). Our densities resemble those of Wässle *et al.*[6], who found 2,233, 3,212, 1,866, 3,254 and 3,005 cells per mm^2.



Extended Data Figure 7 | Alternative contact analysis. Analysis based on summing over BC–SAC pairs rather than averaging as in the main text. **a**, Total BC–SAC contact versus distance from the SAC soma. **b**, Total SAC area within the union of convex hulls of each BC type versus distance. The peak at 80 μm is the location of maximum dendritic branching. The sharp decrease at larger distances is due to thinning and termination of branches. The graphs differ across BC types, which in our sample do not cover exactly the same retinal areas. **c**, Fraction of SAC area in contact with BC types, estimated by dividing contact area (**a**) by SAC area (**b**). This estimate is similar to that of Fig. 4d, but lacks error bars. **d**, Fraction of SAC area contacted by all BC types, the sum of the contact fractions in **c**. Also plotted is the contact predicted by co-stratification, the sum of the curves from Fig. 5b.



Extended Data Figure 8 | Proximity versus contact. Neurons that intermingle may or may not contact each other. **a**, **b**, Type 2 (**a**) and 3a BCs (**b**) contacting SACs. The cells are roughly 24 and 21 μm wide, respectively. **c**, **d**, Other SACs are well within the arborizations of the same two BCs, yet make no contact at all.



Extended Data Figure 9 | Model direction selectivity index (DSI) versus stimulus speed. The graphs are for travelling sine waves of various wavelengths λ (units of Δx). Speed is in units of $\Delta x/\tau$. The preferred speed (horizontal location of each peak) is $\lambda/(2\pi)$. Note that responses are cut off at high speeds by the temporal filters of the model, but the DSI can decay more slowly.

ARTICLE

doi:10.1038/nature13309

c-kit$^+$ cells minimally contribute cardiomyocytes to the heart

Jop H. van Berlo[1,2]*, Onur Kanisicak[1]*, Marjorie Maillet[1], Ronald J. Vagnozzi[1], Jason Karch[1], Suh-Chin J. Lin[1], Ryan C. Middleton[3], Eduardo Marbán[3] & Jeffery D. Molkentin[1,4]

If and how the heart regenerates after an injury event is highly debated. c-kit-expressing cardiac progenitor cells have been reported as the primary source for generation of new myocardium after injury. Here we generated two genetic approaches in mice to examine whether endogenous c-kit$^+$ cells contribute differentiated cardiomyocytes to the heart during development, with ageing or after injury in adulthood. A complementary DNA encoding either Cre recombinase or a tamoxifen-inducible MerCreMer chimaeric protein was targeted to the *Kit* locus in mice and then bred with reporter lines to permanently mark cell lineage. Endogenous c-kit$^+$ cells did produce new cardiomyocytes within the heart, although at a percentage of approximately 0.03 or less, and if a preponderance towards cellular fusion is considered, the percentage falls to below approximately 0.008. By contrast, c-kit$^+$ cells amply generated cardiac endothelial cells. Thus, endogenous c-kit$^+$ cells can generate cardiomyocytes within the heart, although probably at a functionally insignificant level.

The adult mammalian heart was originally proposed to be essentially incapable of renewal after injury or with ageing; although some recent studies have shown that the heart is capable of new cardiomyocyte formation with varying degrees of regenerative potential[1]. The concept that stem cells are the source for cardiomyocyte regeneration arose from initial observations in which bone-marrow-derived c-kit$^+$ haematopoietic stem cells (HSCs) showed restoration of the myocardium after infarction injury when given exogenously[2]. However, subsequent studies demonstrated that HSCs possessed essentially no ability to make cardiomyocytes, calling into question these earlier reports[3,4], at which time the field shifted to a focus on endogenous c-kit$^+$ cardiac progenitor cells (CPCs) residing within the myocardium[5]. Such cells isolated from the rat heart were reported to differentiate into cardiomyocytes, smooth muscle cells and endothelial cells, even after clonal derivation, and when injected into the infarct region they produced substantial new myocardium[6]. Mouse and human c-kit$^+$ CPCs were also isolated and marked, and after injection into an infarcted mouse heart, were shown to generate substantial levels of labelled cardiomyocytes, capillaries and fibroblasts[7]. More recently, resident c-kit$^+$ CPCs were reported to be both necessary and sufficient for complete repair and functional restoration of the myocardium after isoproterenol-induced cardiomyocyte killing, whereas bone-marrow-derived c-kit$^+$ cells had no regenerative effect[8]. However, other studies with adult cardiac resident c-kit$^+$ cells have reported the opposite: that these cells do not possess the ability to generate cardiomyocytes *in vivo*[4,9,10]. To address ongoing controversy, we generated mice in which the *Kit* locus was used for lineage tracing analysis to examine if and how frequently c-kit$^+$ cells generate cardiomyocytes *in vivo*.

c-kit$^+$ contribution to the growing heart

The *Kit* locus was targeted with a cDNA encoding Cre recombinase fused to an internal ribosome entry sequence (IRES) to concurrently express enhanced green fluorescent protein (eGFP)-tagged with a nuclear localization signal (nls) (Fig. 1a). These Kit$^{+/Cre}$ mice were bred to LoxP site-dependent *Rosa26-CAG-loxP-STOP-loxP-eGFP* (R-GFP) reporter mice to irreversibly mark any cell that previously or currently expresses this *Kit* locus (Fig. 1a). Four to twelve weeks after birth, the fidelity of the genetic system was assessed in comparison with known domains of c-kit protein expression, such as melanocytes of the skin, Leydig cells in the testis, interstitial cells of the intestine, lung and wide areas of the spleen, all of which showed eGFP cellular labelling (Fig. 1b and Extended Data Fig. 1a)[11–13]. In bone marrow, 83% of the c-kit-antibody-detected cells were eGFP$^+$ by standard fluorescence-activated cell sorting (FACS) analysis (Fig. 1c), while imaging cytometry analysis detected coincident eGFP$^+$ expression and c-kit immunoreactivity in 88% of the bone marrow cells and 76% of the non-myocyte fraction from the heart (Fig. 1d, e). To further verify the specificity of the *Kit*-Cre allele we examined real-time eGFPnls expression in the heart, ileum and skeletal muscle for co-expression of c-kit protein (antibody), which was always coincident (Fig. 1f, g and Extended Data Fig. 1b, c). In bone marrow, 94% of the eGFP$^+$ cells were Lin$^+$, indicating a high degree of fidelity with the *Kit*-Cre allele (Extended Data Fig. 1d). In the heart c-kit-antibody-positive mononuclear cells were predominantly eGFP$^+$ at 4 weeks of age using the *Kit*$^{+/Cre}$ × R-GFP reporter strategy, whereas in testis recombination was only observed in Leydig cells, of which >80% were eGFP$^+$ (Extended Data Fig. 1e, f). Thus, the specificity of the *Kit*-Cre allele appears identical with known regions of c-kit protein expression *in vivo*.

In an exhaustive search by histological methods across four hearts from Kit$^{+/Cre}$ mice for current eGFPnls expression at 4 weeks of age, no eGFP$^+$ cardiomyocytes or endothelial cells were identified (only mononuclear CPC-like cells were observed), strongly suggesting that the *Kit* locus is not spontaneously activated in differentiated cell types of the heart (Fig. 1f). However, in conjunction with the R-GFP reporter allele for ongoing c-kit lineage tracing, the myocardium showed many eGFP$^+$ differentiated cell types, although cardiomyocytes were very rare (Fig. 1h, i). Even more rarely, areas suggestive of cardiomyocyte clonal expansion were identified (Fig. 1i). No eGFP$^+$ cells were observed in hearts of single R-GFP mice (data not shown). To more rigorously quantify the extent of cardiomyocyte recombination-based labelling,

[1]Department of Pediatrics, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio 45229, USA. [2]Department of Medicine, division of Cardiology, Lillehei Heart Institute, University of Minnesota, Minneapolis, Minnesota 55455, USA. [3]Cedars-Sinai Heart Institute, 8700 Beverly Boulevard, Los Angeles, California 90048, USA. [4]Howard Hughes Medical Institute, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio 45229, USA.
*These authors contributed equally to this work.

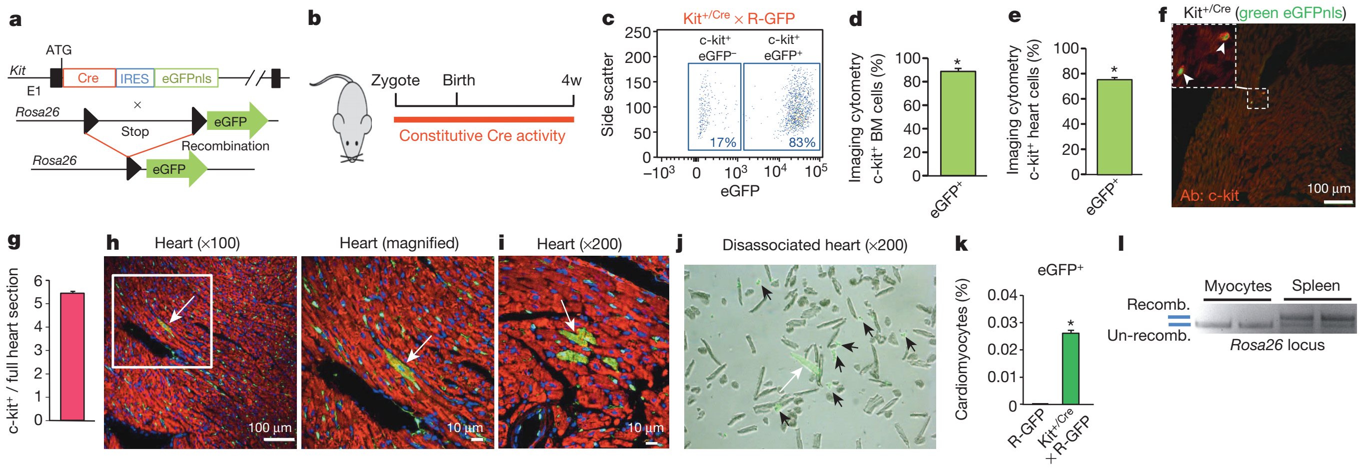
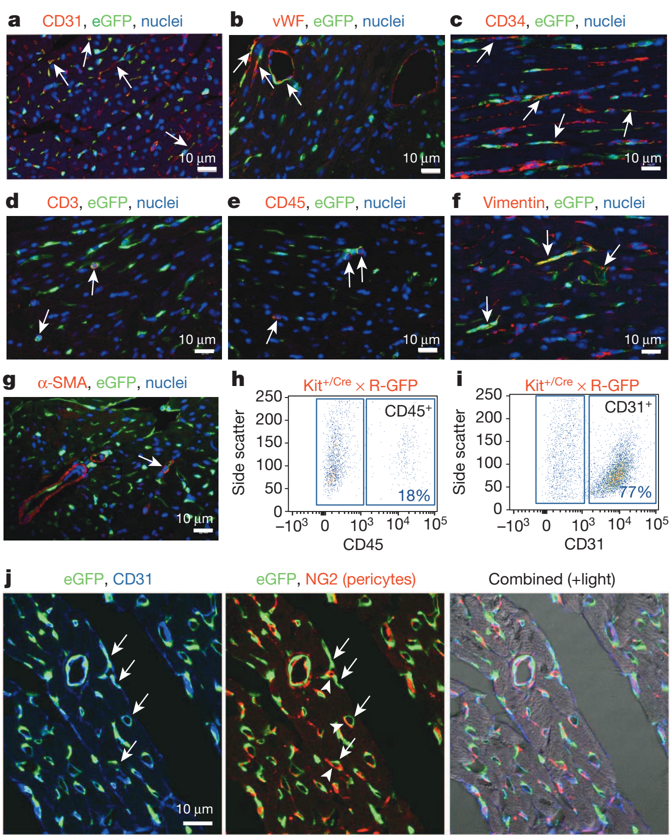


Figure 1 | *Kit*-Cre lineage tracing. a, The *Kit* locus was targeted in mice to express Cre recombinase and eGFP with a nuclear localization sequence (eGFPnls) behind an internal ribosome entry site (IRES). These mice were crossed with *Rosa26* reporter mice (R-GFP) for lineage tracing. **b**, Diagram of mice used for all experimentation in this figure. **c**, Representative FACS plot of bone marrow from $Kit^{+/Cre} \times$ R-GFP mice gated for c-kit antibody, then eGFP fluorescence to reflect recombination of the R-GFP locus (representative of $n = 6$ mice). **d**, Direct imaging cytometry analysis of eGFP expression in bone marrow (averages from $n = 3$ mice, $*P < 0.05$ versus R-GFP). **e**, Same quantitative imaging cytometry analysis as in **d** except the non-myocytes were isolated from hearts of $Kit^{+/Cre} \times$ R-GFP mice (averages from $n = 3$ hearts, $*P < 0.05$ versus R-GFP). **f**, Representative cardiac immunohistochemistry to show current expression from the *Kit*-Cre allele (green, eGFPnls) versus endogenous c-kit protein detected by antibody (Ab, red). The inset box shows two mononuclear c-kit expressing cells. **g**, Quantification of average number of c-kit^{+} cells per longitudinal heart section ($n = 4$ hearts). **h**, Representative histological section at two magnifications (white box) of a $Kit^{+/Cre} \times$ R-GFP mouse heart with desmin antibody in red, eGFP antibody in green, and nuclei in blue. The arrow shows an eGFP^{+} cardiomyocyte. **i**, Representative immunohistological image showing a rare area of cardiomyocyte clonal expansion (arrow) ($n = 6$ hearts analysed). **j**, Image of cells disassociated from the hearts of $Kit^{+/Cre} \times$ R-GFP mice ($n = 3$ hearts analysed). White arrow shows a rare eGFP fluorescing cardiomyocyte, black arrowheads show eGFP fluorescent non-myocytes. **k**, Quantification of eGFP^{+} fluorescent cardiomyocytes (81 from 303,264 total cardiomyocytes, 3 hearts, $*P < 0.05$ versus R-GFP). **l**, DNA electrophoresis after PCR showing Cre-mediated *Rosa26* locus recombination in semi-purified cardiomyocytes and spleens ($n = 2$ $Kit^{+/Cre} \times$ R-GFP mice). All error bars represent s.e.m.

hearts were disassociated and eGFP^{+} cells were directly counted (Fig. 1j), revealing a level of 0.027% myocytes from the c-kit lineage (Fig. 1k). This low percentage was confirmed by PCR analysis for DNA recombination at the *Rosa26* locus from purified cardiomyocytes versus spleen (Fig. 1l).



c-kit^{+} non-myocyte lineage analysis

Hearts of $Kit^{+/Cre} \times$ R-GFP mice at 4–12 weeks of age were further examined to identify the remaining eGFP^{+} non-myocytes. Examples of eGFP labelling co-incident with fibroblasts (vimentin co-labelling), endothelial cells (CD31, CD34, von Willebrand factor (vWF)), immune cells (CD3, CD45) and, rarely, smooth muscle α-actin (α-SMA)-expressing cells, were identified, although the most prevalent co-localizations were with CD31-, CD45- or CD34-positive cells (Fig. 2a–g). Indeed, using a cocktail of antibodies for CD31, CD45, CD34 and CD3, versus sarcomeric α-actin, we were able to account for almost all eGFP^{+} non-myocytes in the hearts of adult $Kit^{+/Cre} \times$ R-GFP mice, either when analysed from histological sections or as dissociated individual cells (Extended Data Fig. 2a–c). FACS analysis showed that 18% and 77% of the total eGFP^{+} non-myocytes in the heart were CD45 or CD31 positive, respectively (Fig. 2h, i). Confocal microscopy analysis showed exact co-localization between eGFP^{+} cells in the heart and CD31 protein expression, but not with NG2 staining for pericytes (Fig. 2j).

We also collected $Kit^{+/Cre} \times$ R-GFP mice at birth (postnatal day (P)0) to analyse the contributions of c-kit^{+} cells to the heart during embryonic and fetal development (Extended Data Fig. 3a). Control histological sections from the ileum and lung showed the expected distribution of c-kit^{+} cells (Extended Data Fig. 3b), and the heart also showed numerous eGFP^{+} cells throughout (Extended Data Fig. 3c). Immunohistochemical

Figure 2 | Analysis of cardiac cells from $Kit^{+/Cre} \times$ R-GFP mice. a–g, Representative immunofluorescent images of heart histological sections from $Kit^{+/Cre} \times$ R-GFP mice at 4 weeks of age stained with eGFP antibody (green), nuclei in blue and either CD31, vWF, CD34, CD3, CD45, vimentin or α-SMA in red. Arrows show cells with overlap in staining ($n = 3$ hearts examined). **h, i**, FACS plot showing lineage markers of heart-isolated c-kit-derived eGFP^{+} cells for CD45 (**h**) and CD31 (**i**) (representative of $n = 6$ for CD45 at 4 weeks of age, and $n = 3$ for CD31 at 12 weeks of age). **j**, Representative immunofluorescent image from heart histological section of a $Kit^{+/Cre} \times$ R-GFP mouse at 4 weeks for eGFP fluorescence (green), CD31 antibody staining (blue) and NG2 antibody staining (red). Right panel shows composite with transmitted light ($n = 2$ hearts analysed).

analysis of the P0 heart with a sarcomeric cardiomyocyte marker showed that nearly all of the eGFP$^+$ cells were non-myocytes, although definable cardiomyocytes were clearly present at very low levels, including rare areas of cardiomyocyte clonal expansion (Extended Data Fig. 3d–g).

c-kit$^+$ lineage tracing in adult heart

To specifically address the question of new cardiomyocyte formation within the adult heart, we generated a mouse model in which the tamoxifen-inducible MerCreMer protein was targeted to the *Kit* locus (Kit$^{+/MCM}$), followed by cross breeding with the R-GFP reporter line (Fig. 3a). To verify the fidelity of this system, Kit$^{+/MCM}$ × R-GFP mice were given tamoxifen during postnatal maturation for approximately 4 weeks followed by collection of tissues with known sites of c-kit expression (Extended Data Fig. 4a). Kit$^{+/MCM}$ × R-GFP mice showed ≈70% overlap in recombination-dependent eGFP expression and endogenous c-kit protein in Leydig cells of the testis (Extended Data Fig. 4b). Importantly, no eGFP$^+$ cells were observed in the absence of tamoxifen at any age examined or after myocardial infarction injury, demonstrating that the MerCreMer system does not 'leak' (Extended Data Fig. 4c). Kit$^{+/MCM}$ × R-GFP mice were also given tamoxifen from day 1 through 6 months of age for continuous labelling (Fig. 3b), which produced eGFP expression in greater than 60% of bone marrow cells, but again no signal in the absence of tamoxifen (Fig. 3c–e). Histological analysis of the heart after 6 months of labelling showed rare examples of eGFP$^+$ adult cardiomyocytes and a relatively large number of non-myocytes (Fig. 3f, g). Careful analysis of the non-myocyte fraction in these hearts showed fibroblasts (rarely), smooth muscle cells (rarely), endothelial cells and immune cells, with the majority again being CD31$^+$ (Extended Data Fig. 5a–g). Myocardial infarction injury also doubled the number of CD31 cells that were eGFP$^+$ in the adult heart with 8 weeks of prior tamoxifen labelling (Extended Data Fig. 5h). We also conducted c-kit lineage labelling from 6–12 weeks of age, just after the postnatal developmental period (Fig. 3h). Upon disassociation of these hearts we observed 0.0055% eGFP$^+$ adult cardiomyocytes (Fig. 3i, j), confirmed as extremely low by PCR and quantitative PCR (qPCR) for *Rosa26* locus recombination (Extended Data Fig. 6a–c).

Cardiac injury increases cellular turnover in the heart, hence we subjected Kit$^{+/MCM}$ × R-GFP mice to myocardial infarction at 10 weeks of age during a 6-week tamoxifen-labelling protocol (Fig. 3k and Extended Data Fig. 6d–f). The percentage of eGFP$^+$ cardiomyocytes increased to

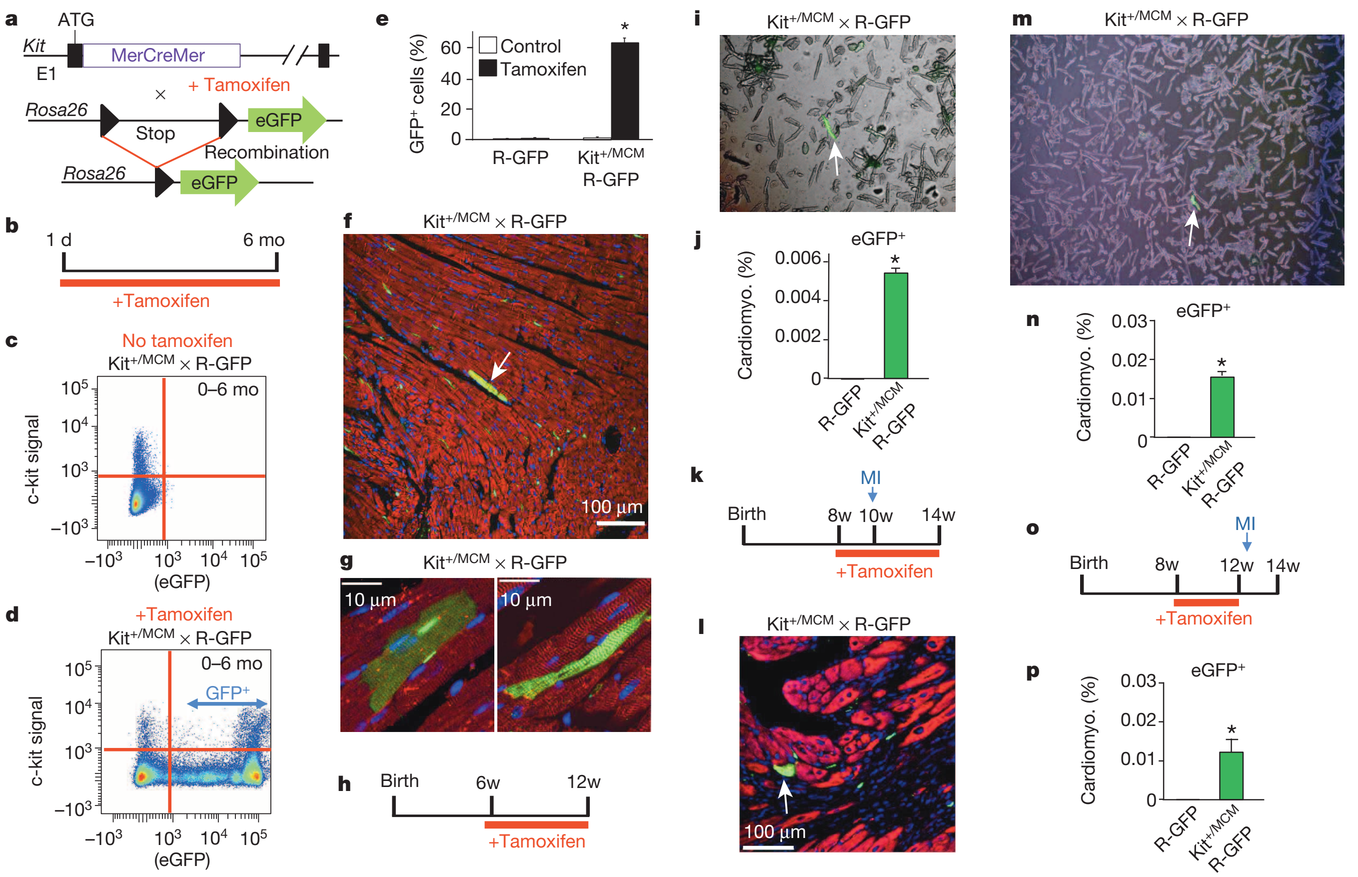


Figure 3 | Inducible Cre expression from the *Kit* locus shows limited adult cardiomyocyte formation. **a**, Genetic cross between Kit$^{+/MCM}$ and R-GFP reporter mice to lineage trace c-kit-expressing cells when tamoxifen is present. **b**, Schematic showing tamoxifen treatment between day 1 and 6 months of age (panels **c**–**g**). **c**, **d**, Representative FACS plots with c-kit antibody versus eGFP from bone marrow of Kit$^{+/MCM}$ × R-GFP mice without (**c**) or with (**d**) tamoxifen. **e**, FACS quantification of eGFP$^+$ cells from bone marrow of these mice (average from $n = 2$ mice for R-GFP and $n = 4$ for Kit$^{+/MCM}$ × R-GFP). $*P < 0.05$ versus R-GFP. **f**, **g**, Representative heart sections from Kit$^{+/MCM}$ × R-GFP mice showing c-kit$^+$ lineage cells in green and cardiomyocytes in red (desmin antibody). White arrow indicates eGFP$^+$ adult cardiomyocyte ($n = 3$ mice analysed). **h**–**j**, Tamoxifen treatment of Kit$^{+/MCM}$ × R-GFP mice between 6–12 weeks of age followed by disassociation of cells from the hearts of these mice in **h** (white arrow in **i** shows rare cardiomyocyte) that is quantified in **j** (127,284 cardiomyocytes across two hearts, 7 were eGFP$^+$, $*P < 0.05$ versus R-GFP). **k**–**n**, Tamoxifen treatment of Kit$^{+/MCM}$ × R-GFP mice between 8 and 14 weeks of age with myocardial infarction (MI) on week 10 ($n = 3$ mice analysed). **l**, Immunohistological heart section for desmin (red) and eGFP (green) with nuclei in blue (arrow shows a cardiomyocyte from the c-kit$^+$ lineage). **m**, **n**, Disassociated cardiomyocytes show rare but definitive myocyte labelling (white arrow), which was quantified in **n** (225,760 cardiomyocytes from 2 myocardial infarction-injured hearts, 37 were eGFP$^+$, $*P < 0.05$ versus R-GFP). **o**, **p**, Tamoxifen treatment between 8 and 12 weeks of age with myocardial infarction injury occurring 3 days after tamoxifen cessation. **p**, Average number of eGFP$^+$ cardiomyocytes from histological sections taken across the entire heart ($n = 2$ hearts, >50 sections each) $*P < 0.05$ versus R-GFP. All error bars represent s.e.m.

0.016% within the heart, with more being localized to the infarct border zone (Fig. 3l–n). c-kit$^+$ lineage cells within the heart were also prelabelled by giving tamoxifen only before myocardial infarction injury, which again showed a very low percentage of eGFP$^+$ cardiomyocytes (Fig. 3o, p). Percentages of eGFP$^+$ cardiomyocytes in the heart during 4 weeks of isoproterenol infusion-induced injury were 0.007% (Extended Data Fig. 7a–c). These astonishingly low values of cardiomyocyte formation were independently verified using blinded heart histological sections from Kit$^{+/MCM}$ × R-GFP mice sent to an outside academic laboratory (Extended Data Fig. 8a–c).

Finally, we also cultured total non-myocytes from the hearts of young adult Kit$^{+/Cre}$ × R-GFP mice in the presence of dexamethasone as a means of pushing c-kit$^+$ cells with progenitor-like activity towards the cardiomyocyte lineage (Extended Data Fig. 9). The data show that eGFP$^+$, *Kit*-Cre allele expressing cells are fully capable of inducing expression of the cardiac markers GATA4, α-actinin and troponin T, suggestive of partial differentiation towards the cardiomyocyte lineage (sarcomeres were not observed).

c-kit$^+$ cells fuse in the heart

Hearts from Kit$^{+/MCM}$ × R-GFP mice showed the presence of cells from blood lineages (CD3, CD45, CD34), which are known to have fusigenic activity with resident parenchymal cells[3,14–18]. To examine fusion we used a genetic strategy that constitutively expresses a membrane-targeted fluorescent tdTomato protein from the *Rosa26* locus. Upon Cre-mediated recombination, tdTomato fluorescence is lost and a membrane-targeted eGFP becomes expressed (abbreviated mT/mG for membrane-targeted tdTomato and eGFP, respectively) (Fig. 4a). If cells fuse, both signals would be present but a *de novo* cardiomyocyte from a c-kit$^+$ lineage cell would be only green. Experimentally, Kit$^{+/MCM}$ × mT/mG mice were given tamoxifen for 2 weeks (8–10 weeks of age) then 3 days later myocardial infarctions, followed by collection at 1, 2 and 4 weeks thereafter (Fig. 4b). Control mice were collected before myocardial infarction but after tamoxifen (time 0). Percentages of total cardiomyocyte membrane-eGFP labelling, whether from fusion or not, were approximately 0.01% at all three time points after myocardial infarction (Fig. 4c). Although some *de novo* cardiomyocytes were identified in the heart (eGFP only), the majority (80–88%) retained the membrane-tdTomato label indicating that these cells probably arose by fusion (Fig. 4d–f). Thus, c-kit$^+$ lineage cells can generate cardiomyocytes in the heart, although at ≈5-fold lower values than initially predicted.

Kit-Cre locus is not ectopically induced

One concern with the *Kit* allele-based lineage tracing approach is that if this locus ever becomes activated ectopically in a cardiomyocyte, it would be wrongly ascribed as having come from a c-kit$^+$ cell. It was previously shown that knockdown of the *Kit* gene in naturally occurring *w* and *wv* mutant mice results in defective progenitor cell activity in many tissues[19–22]. Indeed, hearts from *Kit*$^{w/wv}$ mice showed a marked reduction in resident mononuclear c-kit$^+$ cells and progenitor activity[23]. Hence, *Kit* null mice should lack the ability to generate eGFP$^+$ cardiomyocytes in the heart if they indeed arise from c-kit$^+$ cells with progenitor-like activity, instead of having arisen from ectopic *Kit* allele induction in a rare population of differentiated cardiomyocytes.

Kit null mice were generated by placing the *Kit*-Cre allele over the *Kit*-MerCreMer allele. Although these mice die at birth, viable nulls at embryonic days 16.5 and 18.5 were identified and examined (Fig. 4g–i). Fourteen total eGFP$^+$ cardiomyocytes were counted from four Kit$^{+/Cre}$

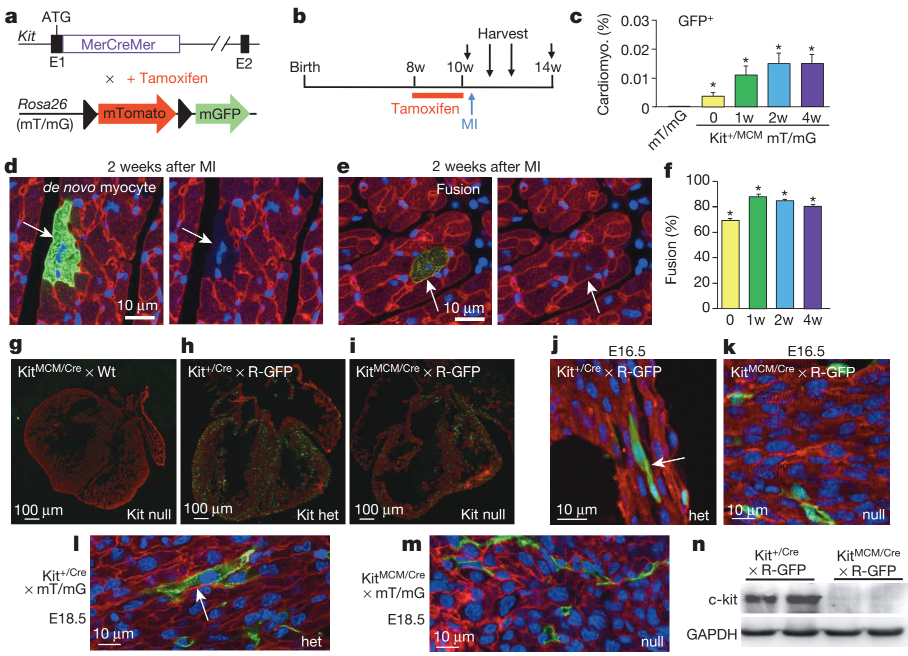


Figure 4 | Assessment of fusion versus *de novo* cardiomyocyte formation in the heart. **a**, Genetic strategy in which Kit$^{+/MCM}$ mice were crossed with *Rosa26* targeted mice containing the membrane targeted tdTomato/eGFP (mT/mG) reporter. **b–f**, Tamoxifen was given to Kit$^{+/MCM}$ × mT/mG mice between 8 and 10 weeks, followed 3 days later by myocardial infarction injury. **c**, Quantification across >50 histological sections of all eGFP$^+$-expressing cardiomyocytes (averages) before myocardial infarction ($n = 4$ hearts) and 1 ($n = 4$ hearts), 2 ($n = 5$ hearts) and 4 ($n = 3$ hearts) weeks after myocardial infarction injury. Error bars represent s.e.m., *$P < 0.05$ versus mT/mG. **d**, Example of a c-kit-lineage-derived *de novo* cardiomyocyte in which membrane-eGFP (green, left) is expressed and tdTomato fluorescence (red, right) is lost. **e**, Example of eGFP$^+$ cardiomyocyte (green) that still contains endogenous membrane-tdTomato fluorescence (red), indicating fusion. Nuclei are stained blue. **f**, Quantification of fusion percentage. *$P < 0.05$ versus mT/mG. **g–i**, Representative immunohistological images from embryonic (E) day 16.5 mouse hearts that are either Kit$^{+/Cre}$ × R-GFP (het, $n = 4$ (**h**)), Kit$^{MCM/Cre}$ × R-GFP (null, $n = 2$ (**i**)) or Kit$^{MCM/Cre}$ (null, no reporter, $n=2$ (**g**)). Red staining is α-actinin and green is eGFP. **j**, Higher magnification image from **h**, showing a definitive eGFP$^+$ cardiomyocyte (arrow). **k**, Higher magnification image from **i**, which shows only eGFP$^+$ non-myocytes in *Kit* null hearts. **l**, **m**, Histological heart images from E18.5 Kit$^{+/Cre}$ (het, $n = 1$) and Kit$^{MCM/Cre}$ (null, $n = 1$) embryos containing the mT/mG reporter, again only the heterozygotes show examples of eGFP$^+$ cardiomyocytes (arrow). **n**, Western blot showing loss of c-kit protein in Kit$^{MCM/Cre}$ embryos (nulls) versus heterozygous controls.

$\times$ R-GFP and 1 Kit$^{+/Cre}$ $\times$ mT/mG embryos across 56 histological sections spanning the heart (Fig. 4j, l). However, hearts from two Kit$^{MCM/Cre}$ $\times$ R-GFP and one Kit$^{MCM/Cre}$ $\times$ mT/mG embryos (nulls) showed lower total eGFP$^+$ cells in the heart and no cardiomyocytes across 69 histological sections (Fig. 4i, k, m). Importantly, Kit$^{MCM/Cre}$ embryos showed no c-kit protein expression confirming their null status (Fig. 4n). Taken together, these data indicate that eGFP$^+$ cardiomyocytes that are lineage traced with the *Kit*-Cre allele are not due to inappropriate activation of the *Kit* gene for even a brief period of time in rare existing cardiomyocytes, but rather they either arose by transdifferentiation from c-kit$^+$ lineage precursor cells or by fusion.

Discussion

The original hypothesis that c-kit$^+$ cells have the ability to contribute to the cardiomyocyte compartment of the heart, as well as other cell types, is correct as determined by the lineage tracing technique used here[6]. Indeed, the observation that embryonic and postnatal labelling in the hearts of Kit$^{+/Cre}$ $\times$ R-GFP mice shows definable regions with cardiomyocyte clonal expansion strongly suggests that these c-kit$^+$ cells can make cardiomyocytes *in vivo*. More importantly, loss of the *Kit* gene, which is known to compromise the progenitor and migration activity of c-kit$^+$ cells, completely prevented cardiomyocyte formation from c-kit$^+$ cells. However, throughout development, with ageing or with cardiac injury, the percentage of cardiomyocytes emerging from the c-kit$^+$ lineage was astonishingly low and hence highly unlikely to ever considerably affect cardiac function. The mT/mG detection system also supported the existence of *de novo* cardiomyocyte formation in the adult heart from the c-kit$^+$ lineage but at $\approx$5-fold lower levels than initially quantified owing to prevalent cellular fusion events.

Exogenous c-kit$^+$ cells are currently being used to treat post-myocardial infarction heart failure patients, and early results have shown small, albeit significant, functional improvements in the heart[24]. However, our results suggest that the potential benefit of injecting c-kit$^+$ cells into the hearts of patients is unlikely attributable to new cardiomyocyte formation, hence caution is warranted until the mechanisms in play are better defined, or until we are able to considerably enhance the cardiogenic potential of these cells (see Supplementary Discussion).

METHODS SUMMARY

The *Kit* allele was targeted in SV129 embryonic stem cells to express either Cre recombinase alone or a tamoxifen-inducible Cre recombinase referred to as MerCreMer. Hemizygous targeted mice were crossed with FVB.Cg-*Gt(ROSA)26Sor*$^{tm1(CAG-lacZ,EGFP)Glh}$/J (previously modified by cross-breeding to B6(C3)-Tg(Pgk1-FLPo)10Sykr/J) or B6.129(Cg)-*Gt(ROSA)26Sor*$^{tm4(ACTB-tdTomato,-EGFP)Luo}$/J. Tissues from these mice were subjected to histological analysis and immunohistochemistry at multiple ages and after select treatments. Antibodies used are shown in Supplementary Table 1 (see Methods for more detailed descriptions).

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 2 January; accepted 4 April 2014.

Published online 7 May 2014.

1. Laflamme, M. A. & Murry, C. E. Heart regeneration. *Nature* **473,** 326–335 (2011).
2. Orlic, D. *et al.* Bone marrow cells regenerate infarcted myocardium. *Nature* **410,** 701–705 (2001).
3. Balsam, L. B. *et al.* Haematopoietic stem cells adopt mature haematopoietic fates in ischaemic myocardium. *Nature* **428,** 668–673 (2004).
4. Murry, C. E. *et al.* Haematopoietic stem cells do not transdifferentiate into cardiac myocytes in myocardial infarcts. *Nature* **428,** 664–668 (2004).
5. Anversa, P., Kajstura, J., Rota, M. & Leri, A. Regenerating new heart with stem cells. *J. Clin. Invest.* **123,** 62–70 (2013).
6. Beltrami, A. P. *et al.* Adult cardiac stem cells are multipotent and support myocardial regeneration. *Cell* **114,** 763–776 (2003).
7. Hosoda, T. *et al.* Clonality of mouse and human cardiomyogenesis *in vivo*. *Proc. Natl Acad. Sci. USA* **106,** 17169–17174 (2009).
8. Ellison, G. M. *et al.* Adult c-kit^{pos} cardiac stem cells are necessary and sufficient for functional cardiac regeneration and repair. *Cell* **154,** 827–842 (2013).
9. Jesty, S. A. *et al.* c-kit$^+$ precursors support postinfarction myogenesis in the neonatal, but not adult, heart. *Proc. Natl Acad. Sci. USA* **109,** 13380–13385 (2012).
10. Zaruba, M. M., Soonpaa, M., Reuter, S. & Field, L. J. Cardiomyogenic potential of C-kit$^+$-expressing cells derived from neonatal and adult mouse hearts. *Circulation* **121,** 1992–2000 (2010).
11. Ro, S. *et al.* A model to study the phenotypic changes of interstitial cells of Cajal in gastrointestinal diseases. *Gastroenterology* **138,** 1068–1078 (2010).
12. Okura, M., Maeda, H., Nishikawa, S. & Mizoguchi, M. Effects of monoclonal anti-c-kit antibody (ACK2) on melanocytes in newborn mice. *J. Invest. Dermatol.* **105,** 322–328 (1995).
13. Sandlow, J. I., Feng, H. L., Cohen, M. B. & Sandra, A. Expression of c-KIT and its ligand, stem cell factor, in normal and subfertile human testicular tissue. *J. Androl.* **17,** 403–408 (1996).
14. Alvarez-Dolado, M. *et al.* Fusion of bone-marrow-derived cells with Purkinje neurons, cardiomyocytes and hepatocytes. *Nature* **425,** 968–973 (2003).
15. Matsuura, K. *et al.* Cardiomyocytes fuse with surrounding noncardiomyocytes and reenter the cell cycle. *J. Cell Biol.* **167,** 351–363 (2004).
16. Nygren, J. M. *et al.* Bone marrow-derived hematopoietic cells generate cardiomyocytes at a low frequency through cell fusion, but not transdifferentiation. *Nature Med.* **10,** 494–501 (2004).
17. Oh, H. *et al.* Cardiac progenitor cells from adult myocardium: homing, differentiation, and fusion after infarction. *Proc. Natl Acad. Sci. USA* **100,** 12313–12318 (2003).
18. Terada, N. *et al.* Bone marrow cells adopt the phenotype of other cells by spontaneous cell fusion. *Nature* **416,** 542–545 (2002).
19. Peters, E. M., Tobin, D. J., Botchkareva, N., Maurer, M. & Paus, R. Migration of melanoblasts into the developing murine hair follicle is accompanied by transient c-Kit expression. *J. Histochem. Cytochem.* **50,** 751–766 (2002).
20. Waskow, C., Paul, S., Haller, C., Gassmann, M. & Rodewald, H. R. Viable c-Kit$^{W/W}$ mutants reveal pivotal role for c-kit in the maintenance of lymphopoiesis. *Immunity* **17,** 277–288 (2002).
21. Mauduit, C., Hamamah, S. & Benahmed, M. Stem cell factor/c-kit system in spermatogenesis. *Hum. Reprod. Update* **5,** 535–545 (1999).
22. Ro, S. *et al.* A model to study the phenotypic changes of interstitial cells of Cajal in gastrointestinal diseases. *Gastroenterology* **138,** 1068–1078 (2010).
23. Ye, L. *et al.* Aging Kit mutant mice develop cardiomyopathy. *PLoS ONE* **7,** e33407 (2012).
24. Bolli, R. *et al.* Cardiac stem cells in patients with ischaemic cardiomyopathy (SCIPIO): initial results of a randomised phase 1 trial. *Lancet* **378,** 1847–1857 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by grants from the National Institutes of Health (to J.H.v.B., E.M. and J.D.M.). J.D.M. is an investigator of the Howard Hughes Medical Institute.

Author Contributions J.D.M., J.H.v.B., O.K., M.M., S.-C.J.L. and R.J.V. designed the experiments. S.-C.J.L. designed the *Kit* allele targeting construct and targeted mice. J.H.v.B. and O.K. designed the breeding, performed histological experiments and animal procedures. R.J.V. performed the qPCR assays. M.M. performed immunohistochemistry. J.K. performed cell culture experiments. E.M. and R.C.M. designed and conducted the independent verification immunohistochemistry with blinded samples. J.D.M. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.D.M. (Jeff.Molkentin@cchmc.org).

METHODS

Mice. All experiments involving mice were approved by the Institutional Animal Care and Use Committee (IACUC) at Cincinnati Children's Hospital. No human subjects or human material was used. Targeted *Kit*-Cre-IRES-eGFPnls and *Kit*-MerCreMer mice were generated by standard gene targeting techniques. Homology arms upstream and downstream of the ATG start codon of the *Kit* gene in exon 1 were subcloned into a plasmid backbone containing Amp^r and a diphtheria toxin (DTA) cassette through recombineering. A cDNA encoding either Cre-IRES-eGFPnls (from A. P. McMahon, UCLA) or MerCreMer, as well as an frt site-flanked neomycin selection cassette, were cloned in-frame with the *Kit* ATG start site. Embryonic stem (ES) cells were electroporated with linearized targeting vector. Targeted clones were identified by Southern blot and PCR. ES cell aggregation with 8-cell embryos was used to generate chimaeric mice with the *Kit*-Cre-IRES-eGFPnls construct[25], whereas the *Kit*-MerCreMer mice were generated by blastocyst injection at the Howard Hughes Medical Institute (HHMI) gene-targeting core facility (by C. Guo at HHMI, who also generated the *Kit*-MerCreMer targeting vector and targeted ES cells). Germline transmitting male chimaeras were crossed with *Rosa26*-Flpe females (B6.129S4-*Gt(ROSA)26Sor*$^{tm1(FLP1)Dym}$/RainJ) to delete the neomycin cassette and verified offspring were further backcrossed to C57Bl/6J for five generations. Reporter mice FVB.Cg-*Gt(ROSA)26Sor*$^{tm1(CAG-lacZ,EGFP)Glh}$/J (previously modified by cross-breeding to B6(C3)-Tg(Pgk1-FLPo)10Sykr/J) and B6.129(Cg)-*Gt(ROSA)26Sor*$^{tm4(ACTB-tdTomato,-EGFP)Luo}$/J were purchased from the Jackson Laboratories. *Kit* null mice were generated by breeding male $Kit^{+/Cre}$ with female $Kit^{+/MCM} \times$ R-GFP mice, of which 1:8 embryos are predicted to be $Kit^{MCM/Cre} \times$ R-GFP (nulls, with the reporter). Littermates that were $Kit^{+/Cre} \times$ R-GFP were controls to show the full extent of $eGFP^+$ cardiomyocytes that are possible in the heart. Because *Kit* null mice were not identified at birth in multiple litters, we collected mice from this cross at E16.5 and E18.5, which identified viable *Kit* null embryos. PCR genotyping of *Kit*-Cre-IRES-eGFPnls used the following primers: (wt-Kit-forward: 5′-CTGTAGCAGAGAGAGGAGCT-3′ and Cre-reverse: 5′-CTACACCAGAGACGGAAATCC-3′); *Kit*-MerCreMer (MerCreMer-forward: 5′-CTGAACCGCCCATGATCTATT-3′ and MerCreMer-reverse: 5′-GTGGATGTGGTCCTTCTCTTC-3′); *Kit* (forward: 5′-CTGTAGCAGAGAGAGGAGCT-3′ and reverse: 5′-ACAGAGGGTGCAGTCCTCTT-3′). Mice of various ages were used, as indicated for each experiment. Both male and female mice were used in all experiments.

Animal procedures. Tamoxifen citrate containing chow (Harlan laboratories) was used to activate the inducible MerCreMer protein, thereby inducing Cre recombinase activity. We used the standard 400 mg kg^{-1} chow for all experiments, except for labelling right after birth for which we used 200 mg kg^{-1}. The duration of treatment is indicated within each experiment. Myocardial infarction was induced in mice via permanent surgical ligation of the left coronary artery[26]. In brief, mice (both sexes) were anaesthetized using isoflurane and a left lateral thoracotomy was performed. The left coronary artery was identified and ligated just below the left atrium. After closing the thoracotomy and expelling residual air, the mice were allowed to recover. Two-dimensional M-mode echocardiography was performed on mice anaesthetized with 2% isoflurane, using a Hewlett Packard SONOS 5500 with a 15-MHz transducer. An average of three measurements was taken for each mouse. Group sizes were determined from past experience and based on statistical power calculations, and the number of mice is given in the figure or figure legends. Isoproterenol treatment was given via osmotic minipumps (Alzet) at 60 mg per kg per day (in 1 μM ascorbic acid) for 4 weeks. Mice were either killed by CO_2 asphyxiation followed by cervical dislocation or by deep isoflurane sedation with cervical dislocation followed by excision of the heart. Isolated organs were fixed in 4% paraformaldehyde overnight, then processed for paraffin embedding or fixed for 3 h and immersed in PBS containing 30% sucrose overnight before embedding in OCT (Tissue-Tek) for cryo-sectioning.

Cell isolation. We isolated bone marrow cells by flushing femurs and tibiae with Hanks Balanced Salt Solution (HBSS). In brief, bone marrow was flushed using a 25-gauge needle attached to a syringe containing 10 ml ice-cold HBSS supplemented with 2% fetal calf serum (FCS). Cells were spun at 400*g* for 10 min at 4 °C and pellets were re-suspended in 2% FCS/HBSS. After isolation, cells were kept on ice and further processed for flow cytometry or DNA extraction. Adult cardiomyocytes were isolated by removal of beating hearts from anaesthetized mice and cannulated for retrograde perfusion with modified Tyrode solution (NaCl 120 mM, KCl 14.7 mM, KH_2PO_4 0.6 mM, Na_2HPO_4 0.6 mM, $MgSO_4$ 1.2 mM, HEPES 10 mM, $NaHCO_3$ 4.6 mM, taurine 30 mM, glucose 5.5 mM, butanedione monoxime (BDM) 10 mM, pH 7.40) supplemented with Liberase TH (Roche)[27]. After perfusion, hearts were disassociated into individual cardiomyocytes, calcium was gradually added back and cells were plated on laminin-coated cover slips in modified Tyrode solution supplemented with 1 mg ml^{-1} BDM and immediately counted for $eGFP^+$ cardiomyocytes. After counting, cells were imaged with a Nikon Eclipse TE300 inverted fluorescence microscope. Non-cardiomyocytes from the heart were isolated by retrograde perfusion as described previously[28]. In brief, hearts were perfused with a digestion buffer (NaCl 126 mM, KCl 4.4 mM, $MgCl_2$ 5 mM, Na pyruvate 5 mM, NaH_2PO_4 5 mM, creatine 5 mM, HEPES 5 mM, glucose 22 mM, taurine 20 mM) containing 15 μM $CaCl_2$, collagenase type 2 (Worthington, 274 U ml^{-1}) and Protease XIV (Sigma-Aldrich, 0.57 U ml^{-1}). Cardiomyocytes were eliminated by two serial centrifugations at 10*g* for 5 min at 4 °C and the non-cardiomyocyte cell fraction was collected after a final centrifugation at 500*g* for 10 min at 4 °C.

Flow cytometry. Flow cytometry was performed on bone marrow and non-myocyte heart fractions using a BD FACSCanto II running FACSDiva software with the following configuration: 405-nm laser for Alexa 405, 633 nm for APC (allophycocyanin) and 488 nm for GFP. Voltages were determined using single-stain and fluorescence minus one (FMO) controls. Analysis was performed using FlowJo vX. Haematopoietic lineage committed bone marrow cells were identified and negatively gated using a panel of mouse antibodies (CD3e, CD11b, CD45R/B220, Ly6G and Ly-6C and TER-119; collectively Lin^-). c-kit^+ cells were identified by antibody labelling and then plotted for endogenous eGFP fluorescence. Alternatively, all bone marrow cells were labelled with c-kit antibody and then plotted for both c-kit positivity and endogenous eGFP fluorescence. Non-myocytes from the heart were first gated for eGFP fluorescence and plotted for CD45 or CD31 positivity using antibodies conjugated to APC for fluorescence intensity separation. Summary of antibodies used is given in Supplementary Table 1.

Multispectral-imaging flow cytometry. Quantitative real-time c-kit and eGFP expression in bone marrow and non-cardiomyocyte cells from the hearts of $Kit^{+/Cre} \times$ R-GFP mice was analysed by ImageStreamX (Amnis), a multispectral flow cytometer combining standard microscopy with flow cytometry. We used the integrated software INSPIRE to run the ImageStreamX. For each experiment, cells were fixed and stained for c-kit antibody reactivity and suspended in 100 μl buffer (cold HBSS with 2% horse serum). Before running the samples, the ImageStreamX was calibrated using SpeedBeads (Amnis). Samples were acquired for unlabelled, single-colour fluorescence controls, then the experimental samples. At least 10,000 experimental cells and 2,000 control cells were acquired for each sample. Images were analysed using IDEAS image-analysis software (Amnis). Summary of antibodies used is given in Supplementary Table 1.

Immunohistochemistry. Please refer to Supplementary Table 1 for all antibody information and dilutions. For paraffin sections, isolated organs were fixed overnight in freshly diluted 4% paraformaldehyde, dehydrated and sectioned at 5 μm. Following citrate antigen retrieval (BioGenex), the sections were blocked for 1 h at room temperature in a blocking solution (PBS with 0.1% cold water fish skin gelatin, 1% bovine serum albumin, 0.1% Tween-20 and 0.05% NaN_3), which was also used to dilute antibodies. For cryosections, isolated organs were fixed for 3 h in freshly diluted 4% paraformaldehyde at 4 °C, rinsed with PBS and cryoprotected in 30% sucrose/PBS overnight before embedding in OCT (Tissue-Tek) and 10-μm cryosections were collected. Cryosections were blocked for 30 min at room temperature in a blocking solution (PBS with 5% goat serum, 2% bovine serum albumin, 0.1% Triton X-100), which was also used to dilute antibodies. Primary antibodies were incubated overnight at 4 °C, secondary antibodies for 2 h at room temperature, washes were performed in PBS. Cryosections were used to visualize native eGFP or tdTomato fluorescence from the different reporters or from the IRES-eGFP cassette built into the *Kit*-Cre allele. 4',6-diamidino-2-phenylindole (DAPI) was used to stain nuclei (usually in blue). Images were acquired on an inverted Nikon A1R confocal microscope using NIS Elements AR 4.13. Some images were further processed in Photoshop or Image J to increase brightness/contrast of individual channels before generating a pseudo-coloured overlay.

Genomic PCR and qPCR. Genomic DNA was prepared from mouse tissues or isolated mouse cardiomyocytes using the DNeasy Blood & Tissue Kit (Qiagen, 69504) per manufacturer's instructions. In brief, cells or tissues were snap-frozen at time of collection then lysed by incubation with proteinase K for 3 h at 56 °C, followed by spin column purification and elution. Samples were treated with RNase A to remove contaminating RNA. PCR was performed to detect recombined and non-recombined *Rosa26* reporter alleles using primers 5′-tctgcttcactctccccatc (forward, against the CAG promoter/enhancer), 5′-gatcagcagcctctgttccaca (forward, against the PGKNeo cassette) and 5′- cgctgaacttgtggccgtttac (reverse, against eGFP). PCR conditions were 96 °C for 2 min to separate strands, followed by 34 cycles of amplification (96 °C for 30 s, 56 °C for 30 s, 72 °C for 30 s) and a 5-min elongation step at 72 °C. PCR products were visualized on an ethidium bromide-stained agarose gel using a UV molecular imager (Bio-Rad). To quantify levels of recombined and non-recombined *Rosa26* alleles in genomic DNA, qPCR was performed using SYBR Green with the same primers used for PCR above (Applied Biosystems), and detection with a Bio-Rad CFx96 thermocycler. Simultaneous reactions using the primers above were performed to detect recombined versus non-recombined alleles.

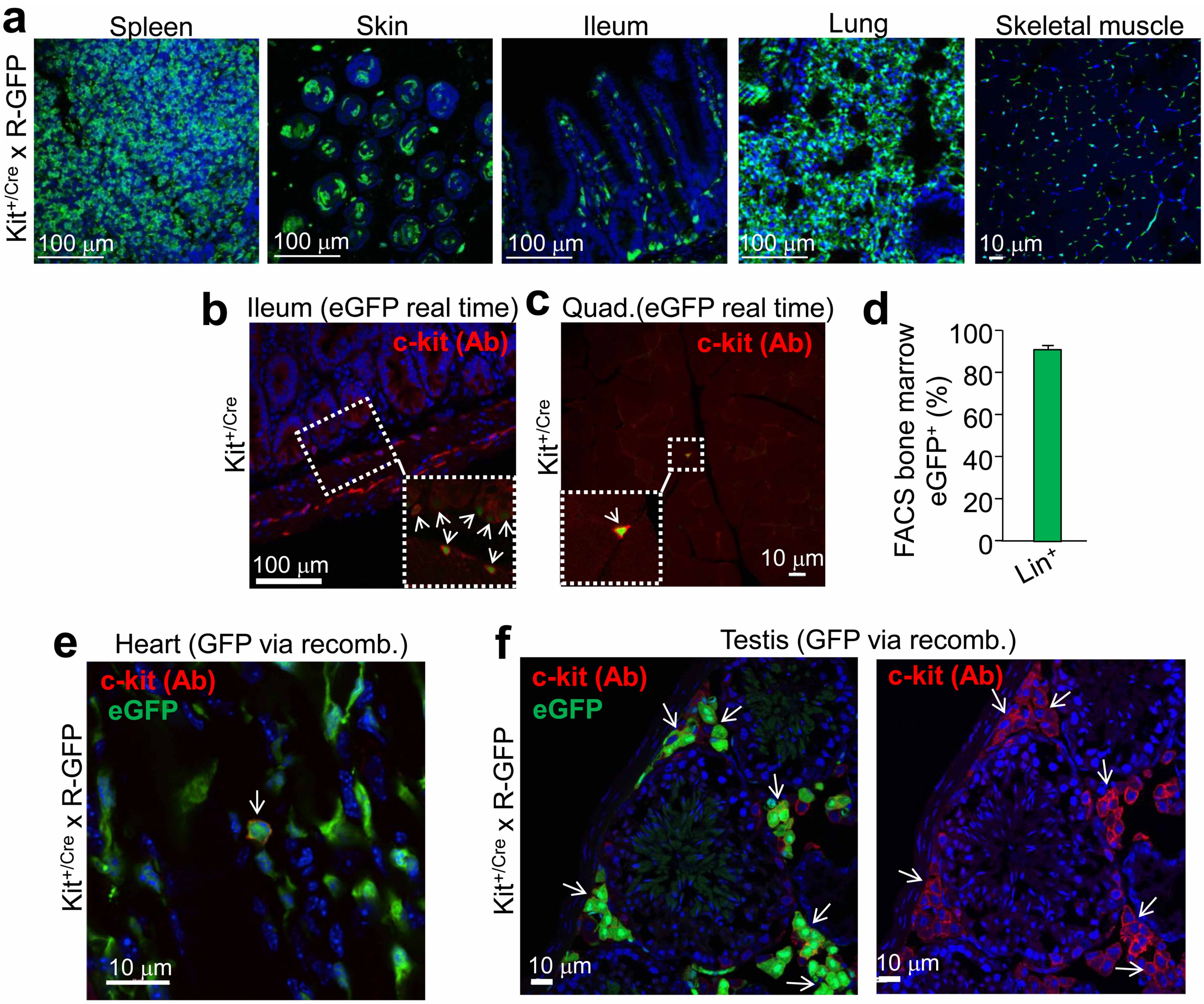
Western blots. Western blotting was performed essentially as described previously[29]. E16.5 embryos were homogenized in RIPA buffer containing protease inhibitor cocktail (Roche) with a dounce homogenizer. Forty micrograms of protein per

sample were resolved on 10% SDS–PAGE gels, transferred onto PVDF membranes, immunoblotted with antibodies for c-kit (R&D Systems AF1356) and GAPDH (Fitzgerald 10R-G109a), and then incubated with the appropriate alkaline phosphate-linked secondary antibodies. The PVDF membranes were visualized by enhanced chemifluorescence (Amersham).

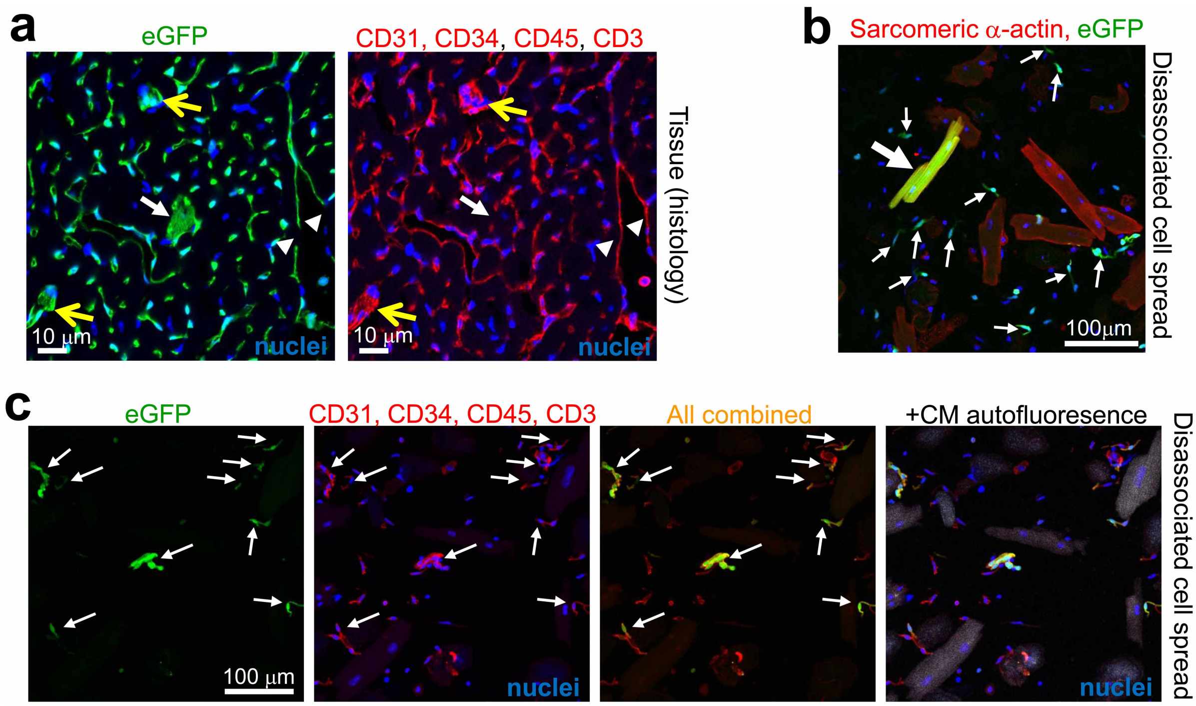
***In vitro* cardiomyocyte differentiation.** The non-cardiomyocyte cell fraction was isolated from a 3-month-old $Kit^{+/Cre} \times$ R-GFP mouse. Cells were plated at a density of 40,000 cells per well on gelatin-coated 6-well tissue culture dishes in DMEM media containing 10% FCS, antibiotics and non-essential amino acids. After 2 days, the cells were washed and treated with 10 nM dexamethasone in DMEM containing 10% FCS to induce differentiation[6]. The media was refreshed every 3 days. After 1 week the cells were fixed with 4% paraformaldehyde and subjected to immunohistochemistry for vimentin, α-actinin, troponin T and GATA4 (antibodies listed in Supplementary Table 1). The cells were then imaged on an inverted Nikon A1R confocal microscope.

Statistics. For studies involving induction of myocardial infarction, group sizes were determined on the basis of previously observed postoperative mortality rates for this procedure. No experimental animals were excluded in any of the analyses. Blinding and randomization were not performed with the exception of the experiments in Extended Data Fig. 8, which were done by two observers blinded to the sample identity. For flow cytometry experiments and direct counting of cardiomyocytes in histological sections or dissociated cardiomyocytes in dishes, two-group comparisons were performed using Student's two-tailed t-test, with $P < 0.05$ considered statistically significant. All error bars throughout the figures are s.e.m. and all represented data are averages.

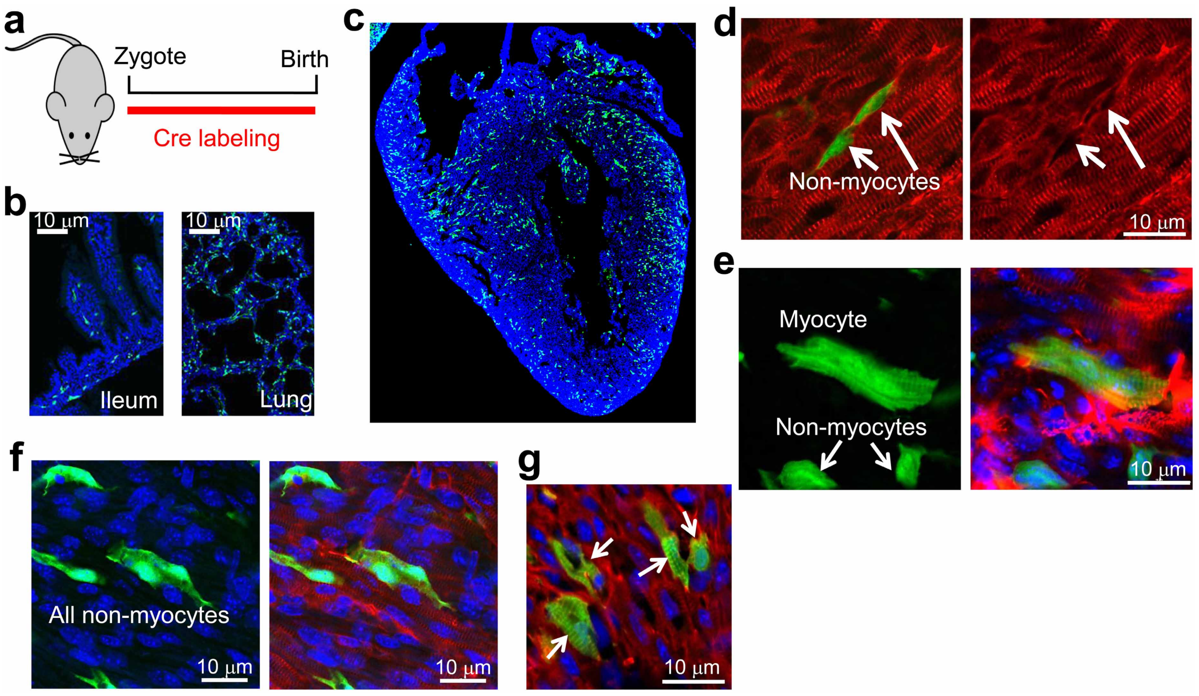
25. Nagy, A. & Rossant, J. in *Gene Targeting: a Practical Approach* (ed. Joyner, A. L.) 2nd edn (Oxford Univ. Press, 2000).
26. Liu, Q., Sargent, M. A., York, A. J. & Molkentin, J. D. ASK1 regulates cardiomyocyte death but not hypertrophy in transgenic mice. *Circ. Res.* **105,** 1110–1117 (2009).
27. Nakayama, H. *et al.* α1G-dependent T-type Ca^{2+} current antagonizes cardiac hypertrophy through a NOS3-dependent mechanism in mice. *J. Clin. Invest.* **119,** 3787–3796 (2009).
28. Hosoda, T. *et al.* Clonality of mouse and human cardiomyogenesis *in vivo*. *Proc. Natl Acad. Sci. USA* **106,** 17169–17174 (2009).
29. Karch, J. *et al.* Bax and Bak function as the outer membrane component of the mitochondrial permeability pore in regulating necrotic cell death in mice. *eLife* **2,** e00772 (2013).



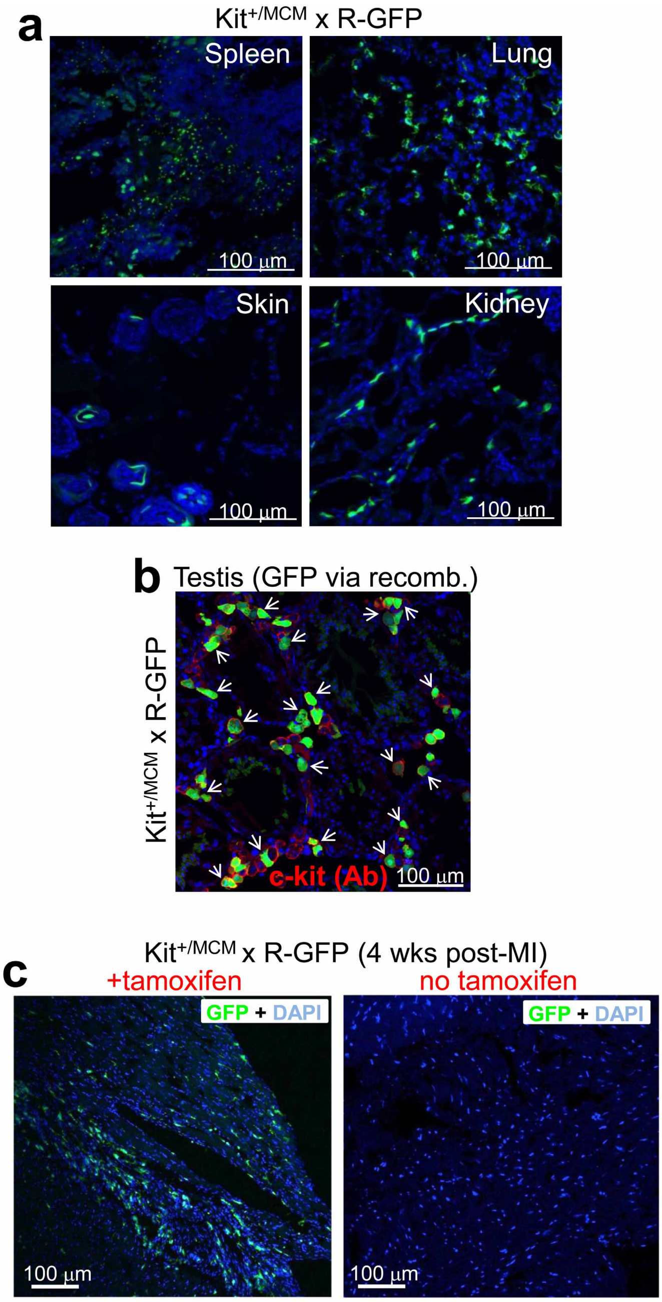
Extended Data Figure 1 | Assessing the fidelity and specificity of the *Kit*-Cre knock-in allele. **a**, Representative histological sections from the indicated tissues of Kit$^{+/Cre}$ × R-GFP mice at 4 weeks of age. Blue is nuclei and green is eGFP. The data show eGFP expression in regions of each tissue that is often characteristic of endogenous c-kit protein expression ($n = 3$ mice). **b**, Immunohistochemistry for endogenous c-kit expression (red) in the mouse ileum at 4 weeks of age from Kit$^{+/Cre}$ mice that contain the IRES-eGFPnls cassette (but without the × R-GFP reporter allele) so that eGFP expression can be monitored in real time. The inset box and arrows show the co-staining with c-kit antibody and eGFP ($n = 3$ mice). **c**, Immunohistochemistry for endogenous c-kit expression (red) in quadriceps muscle of Kit$^{+/Cre}$ mice at 4 weeks of age versus nuclear eGFP (green) from the Kit$^{+/Cre}$ allele ($n = 2$ mice). Although lineage tracing in Kit$^{+/Cre}$ × R-GFP mice, which is cumulative, showed abundant endothelial cells throughout the skeletal muscle (**a**), instantaneous c-kit-expressing cells are rare in skeletal muscle, and when identified, are always mononuclear (inset box). **d**, FACS quantification of bone marrow from Kit$^{+/Cre}$ × R-GFP mice at 4 weeks of age sorted for eGFP expression, of which 94% are positive for the 'lineage' cocktail of differentiation-specific antibodies (averages from $n = 3$ mice, error bars represent s.e.m.). Hence the *Kit*-Cre allele is properly expressed in bone marrow and traces lineages that arise from c-kit^{+} progenitors. **e**, Immunohistochemistry in the hearts of Kit$^{+/Cre}$ × R-GFP mice for endogenous c-kit expression (red) versus all the cells that underwent recombination throughout development and the first 4 weeks of life, shown in green. Although cells that are actively expressing c-kit protein are very rare in the heart (≈5 per heart section), the arrow shows such a cell that is also eGFP^{+} for recombination. All of the currently c-kit-expressing cells identified in the heart were eGFP^{+}, further verifying the fidelity of the *Kit*-Cre allele ($n = 3$ hearts). **f**, Same experiment as in **e** except the testis was examined because of the characteristic pattern of Leydig cells that are known to be actively c-kit-expressing cells ($n = 3$ mice). The data show that greater than 80% of the currently c-kit-antibody-reactive Leydig cells (red outline, better observed in the right panel) are also eGFP^{+} (arrows show clusters of these cells).



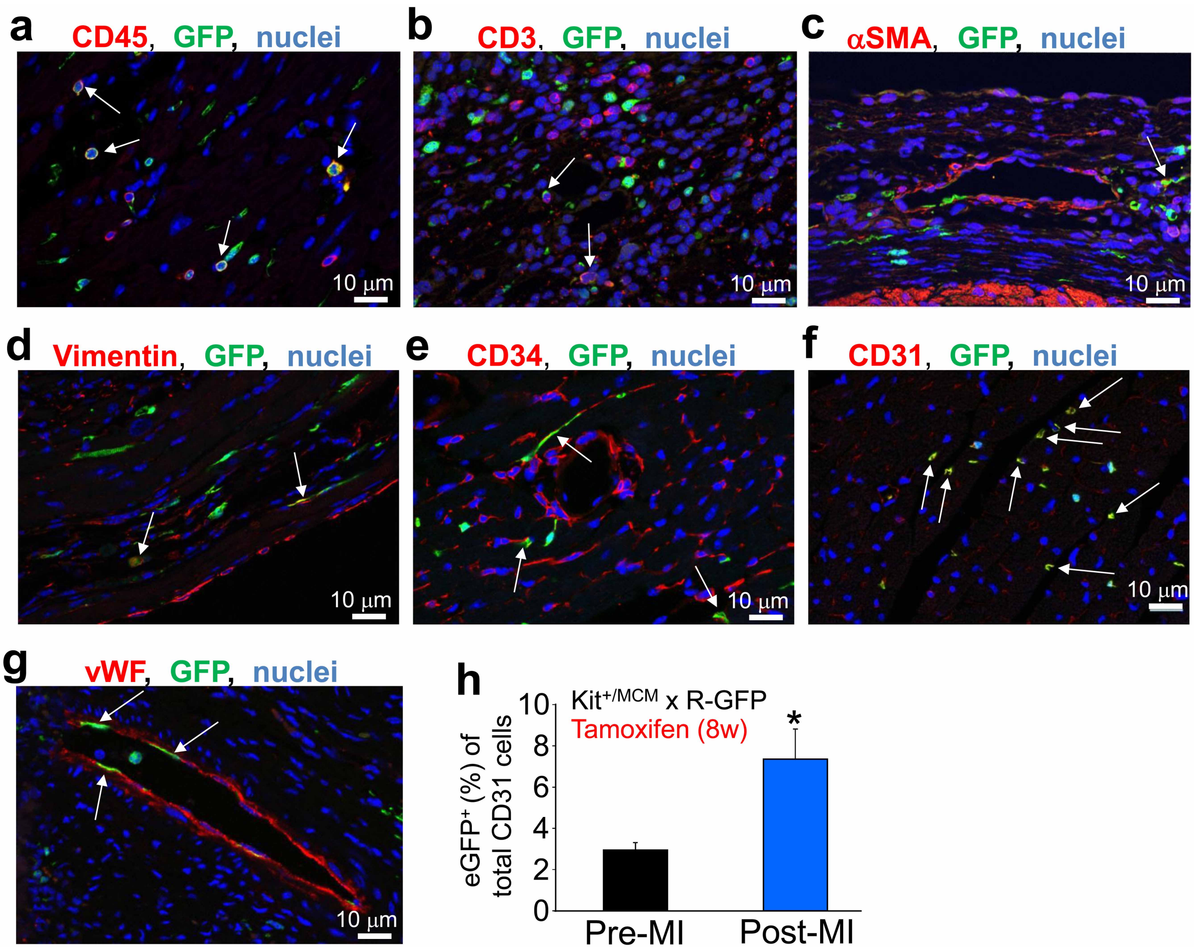
Extended Data Figure 2 | Identification of non-myocytes from the hearts of Kit$^{+/Cre}$ × R-GFP mice. Kit$^{+/Cre}$ × R-GFP mice were collected at 6 weeks of age (constitutive lineage labelling the entire time), although myocardial infarction was performed at week 4 to induce greater vascular remodelling and potentially more c-kit lineage recruitment over the next 2 weeks. **a**, Hearts were then collected at week 6 and subjected to immunohistochemistry with a pool of antibodies for CD31, CD34, CD45 and CD3 in red, whereas the green channel was for eGFP expression from the recombined R-GFP reporter allele due to *Kit*-Cre lineage expression. The white arrowheads show endothelial cells that are not contiguous with the underlying network, although most of the endothelial cells are from the c-kit lineage when the red and green channels are compared. The white arrow shows a cardiomyocyte that lacks red staining, whereas the yellow arrows show two areas with relatively large cells that are eGFP$^+$ and could be mistaken for a cardiomyocyte, although they are also positive for the non-myocyte marker panel of antibodies ($n = 2$ mice). **b**, **c**, Spread of cells isolated from hearts of 8-week-old Kit$^{+/Cre}$ × R-GFP mice at baseline that were subjected to immunocytochemistry for the indicated markers ($n = 3$ hearts). The large white arrow in panel **b** shows an eGFP$^+$ (green) cardiomyocyte that also co-stains with sarcomeric α-actin (red). The smaller arrows show eGFP$^+$ non-myocytes, which in panel **c**, were subject to staining with a cocktail of antibodies again for CD31, CD34, CD45 and CD3 (all in red). This analysis identifies nearly all of the non-myocytes in these cell spreads. The very last image in panel **c** shows a fourth channel with higher gain so that the underlying cardiomyocytes (CMs) autofluoresce (in white) to show the mixed nature of the spread cells. Blue staining depicts nuclei.



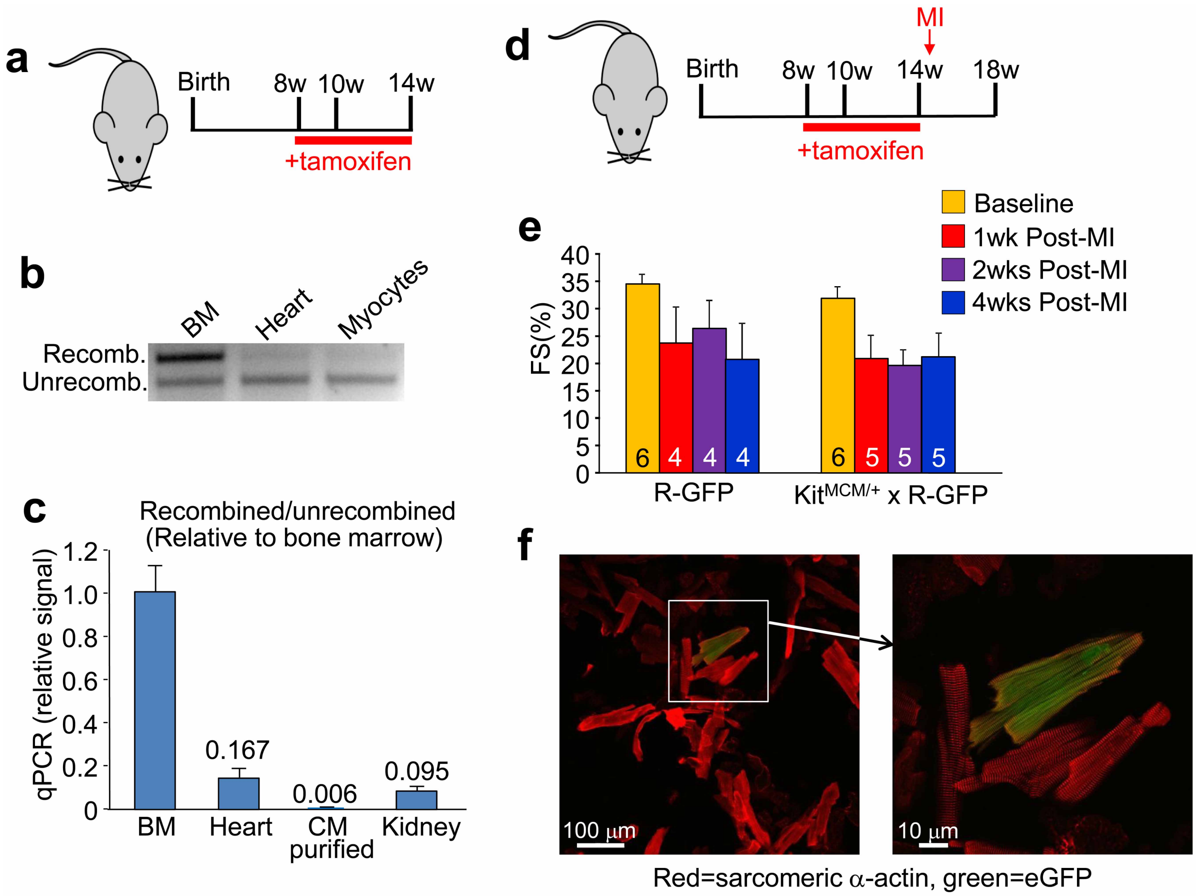
Extended Data Figure 3 | Analysis of c-kit lineage labelling in the heart at P0 (birth). **a**, Diagram of the timing whereby newborn $Kit^{+/Cre}$ × R-GFP mice were analysed for all subsequent experiments in this figure. **b**, Histological sections for eGFP fluorescence (green) from the ileum and lung at P0 showing the characteristic c-kit labelling pattern as observed at other time points or in other studies when antibodies were used. Blue shows nuclei **c**, Histological section for eGFP fluorescence (green) from the heart at P0. Blue shows nuclei and magnification was ×40. **d**, Immunohistochemical tissue section from the P0 heart of $Kit^{+/Cre}$ × R-GFP mice stained with sarcomeric α-actin (red) to show all underlying cardiomyocytes (right panel) or with eGFP expression in green (left panel) as being c-kit-derived. The green cells noted by the arrows are non-myocytes that do not express sarcomeric α-actin. **e**, eGFP expression alone (left) or eGFP with co-staining for cardiomyocytes in red (sarcomeric α-actin) from heart sections at P0 of $Kit^{+/Cre}$ × R-GFP mice ($n = 3$ mice). Blue staining depicts nuclei. The cardiomyocyte that is shown has clear striations in the eGFP staining pattern, whereas the two non-myocytes do not show striated eGFP and also lack sarcomeric α-actin staining. **f**, eGFP expression alone in green (left) with nuclei in blue or eGFP with sarcomeric α-actin co-staining (red) from heart sections at P0 of $Kit^{+/Cre}$ × R-GFP mice. All $eGFP^{+}$ cells shown lack striations and are non-myocytes although the two cells in the centre sit directly on top of cardiomyocytes and could be easily misinterpreted. Great care is needed in scoring myocytes in the P0 heart because they are small and often the same size as $eGFP^{+}$ non-myocytes. **g**, eGFP expression (green) with nuclei in blue and cardiomyocytes identified in red with sarcomeric α-actin antibody from heart histological sections at P0 of $Kit^{+/Cre}$ × R-GFP mice. Here the data show c-kit-lineage-derived cardiomyocytes that appear in a loose cluster (arrows), presumably from a clonal expansion event earlier in development.



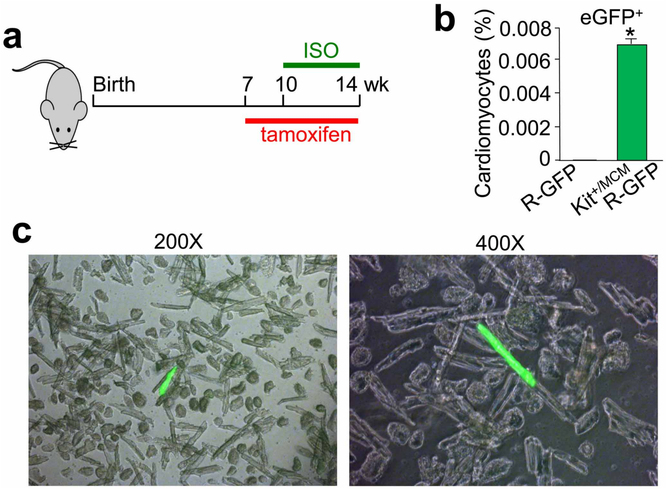
Extended Data Figure 4 | Additional examination of the *Kit*-MerCreMer knock-in allele and its potential leakiness in the absence of tamoxifen. **a**, Histological analysis of eGFP fluorescent cells from the indicated tissues of $Kit^{+/MCM} \times$ R-GFP mice that were given tamoxifen from 2 to 28 days of age and then collected at day 28. Nuclei are shown in blue and green shows $eGFP^{+}$ cells in the expected patterns for known regions of c-kit protein expression, such as the distinct pattern of melanocytes in the skin and widespread expression in the spleen and lungs. **b**, Representative immunohistochemical analysis in the testis of $Kit^{+/MCM} \times$ R-GFP mice for endogenous c-kit expression (red) versus cells that underwent recombination when tamoxifen was given by intraperitoneal injection (2 mg) for five consecutive days (green) ($n = 2$ mice). The data show that most of the cells currently expressing c-kit protein in testis (only Leydig cells react, red surface staining) are also $eGFP^{+}$ (intracellular), indicating that recombination only occurs in c-kit-expressing cells, and in the majority of them. **c**, Representative histological heart sections from $Kit^{+/MCM} \times$ R-GFP mice that were placed on tamoxifen-laden food or vehicle food ($n = 6$ mice per treatment) beginning at 4 weeks of age and then subjected to myocardial infarction injury 4 weeks later, followed by collection 4 weeks after that. In the presence of tamoxifen, histological sections through the myocardial infarction border zone of the heart show widespread $eGFP^{+}$ cells (green) from the c-kit lineage (left panel), whereas in the absence of tamoxifen no $eGFP^{+}$ cells are observed (right panel), indicating the *Kit*-MerCreMer allele does not leak at baseline or after myocardial infarction injury.



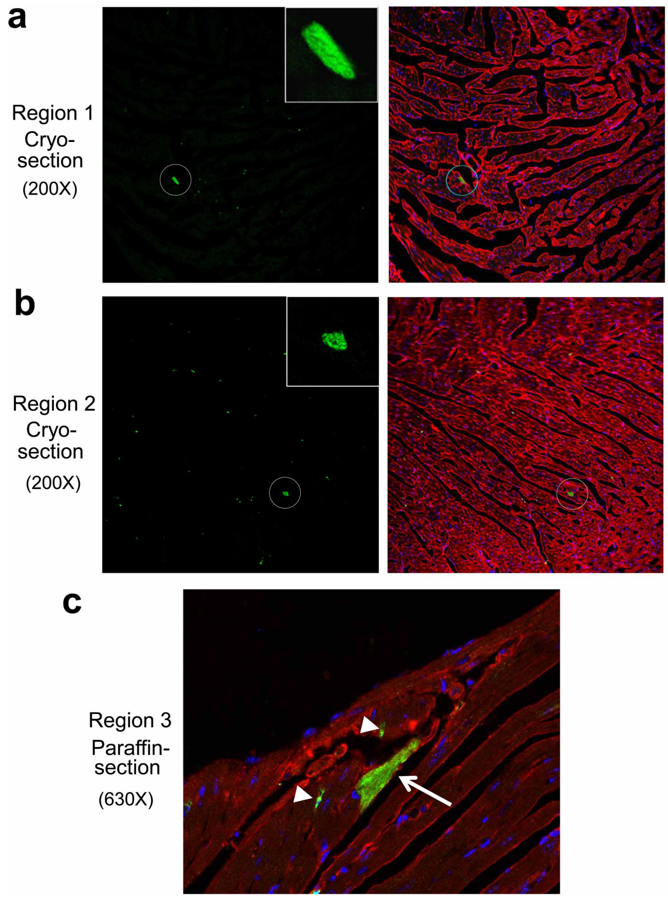
Extended Data Figure 5 | Analysis of eGFP$^+$ non-myocytes in the hearts of Kit$^{+/MCM}$ × R-GFP mice at baseline or after myocardial infarction injury. **a–g**, Tamoxifen was given to Kit$^{+/MCM}$ × R-GFP mice for 1 day–6 months of age (**a**, **e**, **f**) or in mice given tamoxifen and myocardial infarction injury (**b**, **c**, **d**, **g**), followed by collection of the hearts for immunohistochemistry with antibodies for GFP (green), or the indicated antibodies in red: CD45 (**a**), CD3 (**b**), α-SMA (**c**), vimentin (**d**), CD34 (**e**), CD31 (**f**), vWF (**g**). Nuclei are shown in blue. The white arrows show cells with coincident green and red reactivity for each of the markers, although sometimes the red marker is membrane-localized whereas the green (eGFP) is always cytoplasmic. The most overlapping activity with GFP expression was observed for CD31 (endothelial cells), then CD34, followed by CD45 (haematopoietic cells). $n = 2$ Kit$^{+/MCM}$ × R-GFP mice for 1 day–6 months of age; $n = 4$ Kit$^{+/MCM}$ × R-GFP myocardial infarction. **h**, Averages from FACS plots for the CD31 cellular fraction (antibody-detected) in the heart that are also eGFP$^+$ from Kit$^{+/MCM}$ × R-GFP mice (pre-MI, $n = 3$) after 8 weeks of tamoxifen in early adulthood at either baseline or 4 weeks after myocardial infarction injury (post-MI, $n = 3$). The data show about a doubling in the number of CD31 cells that are eGFP$^+$ after myocardial infarction (*$P < 0.05$ versus pre-MI).



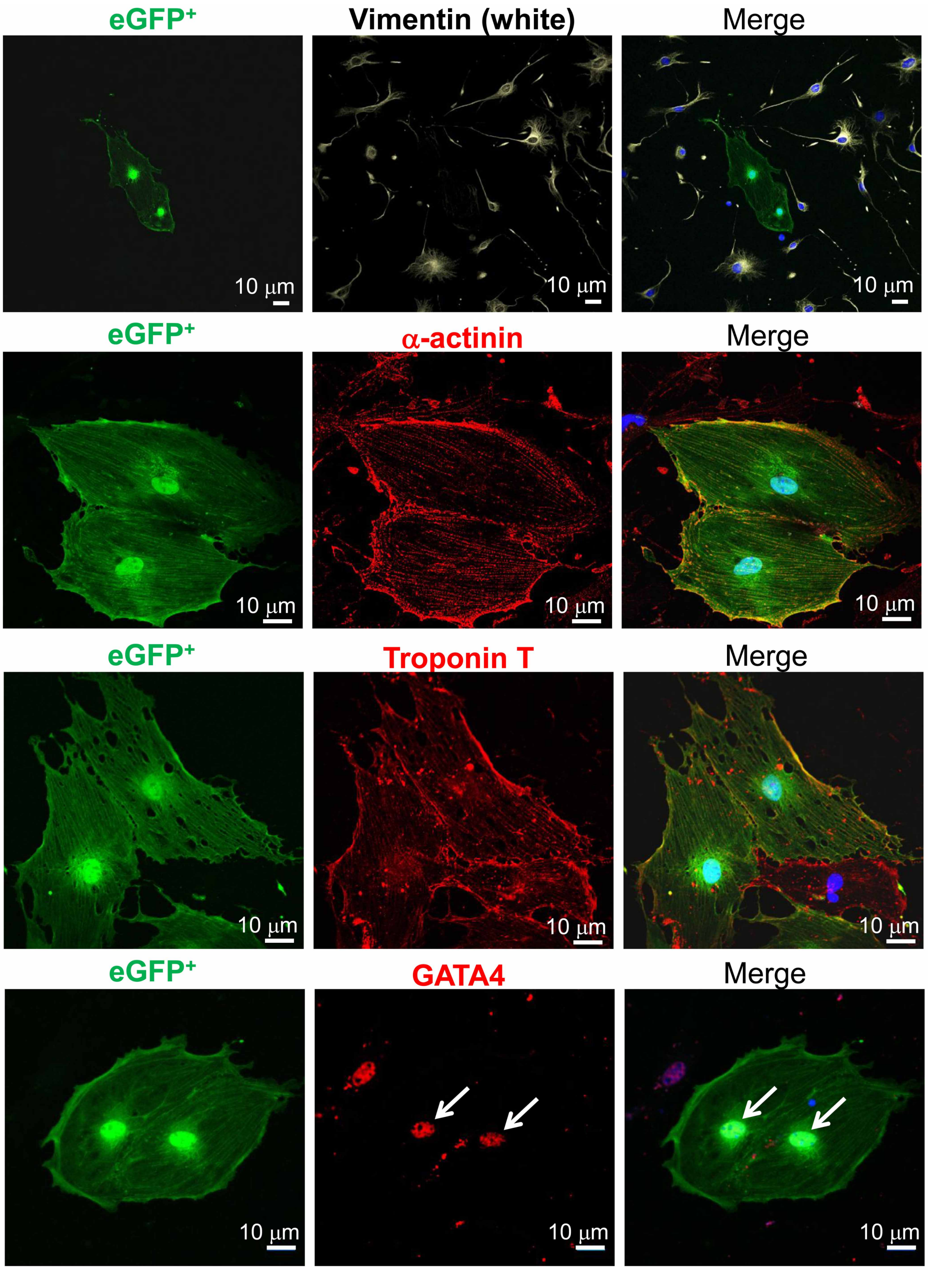
Extended Data Figure 6 | Quantification of Cre activity and DNA recombination in the hearts of Kit$^{+/MCM}$ × R-GFP mice. a, Timeline for tamoxifen administration in Kit$^{+/MCM}$ × R-GFP mice. **b**, PCR from DNA isolated from the bone marrow (BM), whole heart or semi-purified cardiomyocytes after 6 weeks of tamoxifen treatment in Kit$^{+/MCM}$ × R-GFP mice ($n = 2$). Bone marrow shows most of the DNA as having been recombined by Cre, whereas whole heart is just barely discernable, and purified cardiomyocytes show essentially no recombination given the sensitivity constraints of this assay. **c**, qPCR was also run to more sensitively detect and quantify the extent of recombination, which was set relative to the recombination in bone marrow. Semi-purified cardiomyocytes (CM) showed very low rates. Averaged data are shown and error bars are s.e.m. of duplicate technical replicates from $n = 3$ Kit$^{+/MCM}$ × R-GFP mice. **d**, Schematic of the tamoxifen time course and timing of myocardial infarction in Kit$^{+/MCM}$ × R-GFP mice. **e**, Echocardiography measured cardiac fractional shortening (FS%) was assessed in the mice after myocardial infarction, which shows a reduction in cardiac ventricular performance at 1, 2 and 4 weeks after injury. The number of mice analysed is shown in the bars. Error bars represent the s.e.m. Both the control and experimental groups showed an equivalent reduction in cardiac function post-myocardial infarction. **f**, Images of dissociated cardiomyocytes from hearts of Kit$^{+/MCM}$ × R-GFP mice 4 weeks after myocardial infarction, which were fixed and stained for sarcomeric α-actin antibody (red) and eGFP (green) at two different magnifications. One eGFP^{+} cardiomyocyte is shown with sarcomeric patterning of the eGFP fluorescence.



Extended Data Figure 7 | Analysis of eGFP$^+$ myocytes in the hearts of Kit$^{+/MCM}$ × R-GFP mice after isoproterenol infusion-induced injury. **a**, Schematic diagram showing tamoxifen treatment of Kit$^{+/MCM}$ × R-GFP mice between 7 and 14 weeks of age with isoproterenol (ISO) infusion occurring between weeks 10–14. **b**, **c**, Quantification and imaging of disassociated cardiomyocytes (separate images shown at two different magnifications) from the hearts of isoproterenol-injured Kit$^{+/MCM}$ × R-GFP mice, which showed rare but definitive cardiomyocyte labelling. *$P < 0.05$ versus R-GFP, 31 eGFP$^+$ cells of 395,302 counted from two hearts.



Extended Data Figure 8 | Verifying the extent of eGFP$^+$ cardiomyocytes by an independent laboratory from blinded histological heart samples. Unprocessed cryosections and paraffin sections from the hearts of Kit$^{+/MCM}$ × R-GFP mice after 8 weeks of tamoxifen were blinded and sent to the Marbán laboratory along with negative control sections from hearts that should not have staining. **a**, **b**, Two separate images from cryopreserved blocks are shown at ×200 magnification in which the cryosection was processed for eGFP fluorescence (green) and α-actinin antibody (red) to show cardiomyocytes. The data show two regions where a single eGFP$^+$ myocyte is visible in a region with several hundred GFP-negative cardiomyocytes. The single eGFP$^+$ cardiomyocyte is circled and the inset box shows a higher magnification. Sections were also stained for nuclei (blue). In general, approximately 1–2 definitive eGFP$^+$ cardiomyocytes were identified per entire heart section in the Marbán laboratory, a result that is consistent with the approximate numbers of kit lineage-labelled cardiomyocytes observed by us. **c**, Image taken at ×630 magnification from a paraffin-embedded and processed histological section in which both an eGFP antibody (green) and α-actinin antibody (red) was used. Nuclei are shown in blue. The arrow shows a single eGFP$^+$-expressing cardiomyocyte and the arrowheads show eGFP$^+$ non-myocytes.



Extended Data Figure 9 | Assessing cardiomyocyte differentiation markers from total non-myocytes in the heart. Adult cardiac interstitial cells isolated from a Kit$^{+/Cre}$ × R-GFP mouse were treated with dexamethasone for 1 week. Cells were then fixed and subjected to immunocytochemistry for the indicated antibodies. c-kit-lineage-derived cells were green (eGFP^{+}) and showed fluorescence in the cytosol and nucleus. The data show eGFP cells that express markers of differentiated cardiomyocytes such as α-actinin, troponin T and the transcription factor GATA4 (all in red) but not the fibroblast marker vimentin (white), nuclei were stained blue (right panels). These results indicate that eGFP^{+} *Kit*-Cre-expressing cells can generate pre-differentiated cardiomyocytes as well as non-eGFP interstitial cells; hence the cells identified by the *Kit*-Cre (knock-in) reporter strategy are representative of how endogenous c-kit^{+}-expressing cells truly function.

LETTER

doi:10.1038/nature13246

Cepheid variables in the flared outer disk of our galaxy

Michael W. Feast[1,2], John W. Menzies[2], Noriyuki Matsunaga[3] & Patricia A. Whitelock[1,2]

Flaring and warping of the disk of the Milky Way have been inferred from observations of atomic hydrogen[1,2] but stars associated with flaring have not hitherto been reported. In the area beyond the Galactic centre the stars are largely hidden from view by dust, and the kinematic distances of the gas cannot be estimated. Thirty-two possible Cepheid stars (young pulsating variable stars) in the direction of the Galactic bulge were recently identified[3]. With their well-calibrated period–luminosity relationships, Cepheid stars are useful distance indicators[4]. When observations of these stars are made in two colours, so that their distance and reddening can be determined simultaneously, the problems of dust obscuration are minimized. Here we report that five of the candidates are classical Cepheid stars. These five stars are distributed from approximately one to two kiloparsecs above and below the plane of the Galaxy, at radial distances of about 13 to 22 kiloparsecs from the centre. The presence of these relatively young (less than 130 million years old) stars so far from the Galactic plane is puzzling, unless they are in the flared outer disk. If so, they may be associated with the outer molecular arm[5].

We derived the distances for the five Cepheids from near-infrared photometry obtained with the Infrared Survey Facility (IRSF) and we used radial velocities from the Southern African Large Telescope (SALT) to determine the kinematics (see Methods)—both telescopes are at the South African Astronomical Observatory (SAAO), Sutherland, in South Africa. From these data we were able to ascertain the population to which the Cepheids belong. The other 27 Cepheid candidates are either better assigned to a different class (such as anomalous Cepheids) or else their classification as classical Cepheids is uncertain.

Table 1 lists the derived distances and various other parameters for the Cepheids. They are at about the distance and position at which a stream associated with the Sagittarius (Sgr) dwarf galaxy crosses the plane[6], but the low radial velocity (mean heliocentric radial velocity after correction for the effects of stellar pulsation of $V_R = 4 \pm 8\ \mathrm{km\ s^{-1}}$, see Table 1) is completely different from that expected for members of the Sgr dwarf stream (about $150\ \mathrm{km\ s^{-1}}$)[6,7] and the Cepheids are clearly Galactic. They cannot be in the Galactic bulge because their distances from the centre put them far beyond the bulge and the velocity dispersion of the five stars, $16 \pm 5\ \mathrm{km\ s^{-1}}$ (much of which is observational), is much smaller than expected for bulge objects ($>60\ \mathrm{km\ s^{-1}}$)[8]. Furthermore, these short-period Cepheids will be relatively young (about 100 million years (Myr) old), and, although there is a young component, including Cepheids[9], in the innermost regions of the bulge, the bulk of the population is old (about 10 billion years (Gyr) old)[8]. Figure 1 shows the positions of the five stars in comparison to catalogued Cepheids. The various sources of uncertainty for the distances of the Cepheids are discussed in the Methods, but the reddening law and reddening corrections presented the biggest challenge and are the primary contributors to the error bars shown in the figure.

Table 1 | Data for individual Cepheids

OGLE number	l (deg)	b (deg)	D (kpc)	z (kpc)	R (kpc)	V_R ($\mathrm{km\ s^{-1}}$)	ρ ($\mathrm{km\ s^{-1}}$)	P (day)
01	−0.03	2.94	24.7	1.3	16.2	−12	−3	2.598
02	4.57	4.85	23.2	2.0	14.7	+31	50	2.026
03	4.35	2.89	22.1	1.1	13.6	+5	24	1.236
05	5.38	2.34	22.3	0.9	13.8	+7	28	3.796
32	6.89	−3.89	30.4	−2.1	22.0	−10	15	3.736

OGLE numbers are prefixed by 'OGLE-BLG-CEP-'. l and b are the Galactic coordinates. D is the distance from the Sun (D is uncertain to less than about 2 kpc), z is the distance from the Galactic plane and R is the perpendicular distance from the axis of Galactic rotation (assuming the distance from the Sun to the Galactic centre is 8.5 kpc). V_R is the measured heliocentric radial velocity corrected for pulsation ($\pm 15\ \mathrm{km\ s^{-1}}$). ρ is the radial velocity after correction for solar motion, Galactic rotation and the effects of stellar pulsation. P is the pulsation period.

There is almost no information on gas or stars in the Galactic disk immediately behind (Galactic longitude $l \pm 15^\circ$) the Galactic centre. The atomic hydrogen observations[2] on either side of the centre, but away from the central region itself, suggest that the gaseous disk of the Milky Way at $l \approx 0$ is not warped but shows a marked flaring at Galactocentric radii (R, the distance from a star to the centre of the Galaxy) of 15 kiloparsecs (kpc) and more; we note that the details are model dependent. The thickness of the gaseous disk[1,2] increases from 60 parsecs (pc) half-width at half-maximum (HWHM) at $R = 4$ kpc to 2.7 kpc at $R = 30$ kpc and, especially at positive Galactic longitudes[1], there is a marked increase from about 0.4 kpc at $R = 15$ kpc to about 1.0 kpc at $R = 20$ kpc.

Therefore we found the Cepheids at exactly the distance predicted for this increase in disk thickness, as can be seen in Fig. 1. The absence of Cepheids nearer the Sun is consistent with the lower HWHM in these regions, whereas the absence of more distant Cepheids is partly due to the decreasing density at larger distances from the centre and partly the consequence of the Optical Gravitational Lensing Experiment (OGLE) observational cut-off. So the relatively narrow range of distances is consistent with our hypothesis that these stars are in the flared disk. In the Methods we also show that the numbers of Cepheids observed is consistent with expectations from a flared disk.

Cepheids are usually associated with spiral arms and the distances of these five are similar to that expected for the far outer molecular spiral arm of the Galaxy[5] where it passes behind the central region of the Galaxy; the HWHM of this arm may be only about 0.6 kpc, in which case the Cepheids would be on its periphery. However, we note that distances and thickness computed for this arm depend sensitively on the model adopted and are therefore uncertain[2,5].

It is instructive to examine why the outer regions of a galactic disk flare. In the inner parts of a galactic disk the gravitational force $k(z)$ at height z perpendicular to the galactic plane is dominated by the strong concentration of stars there. As we move to greater galactocentric radii, however, the concentration of stars drops dramatically, $k(z)$ decreases and is increasingly dominated by the effects of dark matter. The flaring of the gas layer in the outer parts of our own and other galaxies has been attributed to this, and observations can in principle be used to study the distribution of dark matter in the halo of galaxies[11]. Studies of the flaring of H I gas in our Galaxy[1] suggest that in addition to an isothermal dark halo of $1.8 \times 10^{12} M_\odot$ where $M_\odot$ is the mass of the Sun, there is a self-gravitating exponential dark-matter disk (1.8–$2.4 \times 10^{11} M_\odot$) as well as a dark-matter ring (13 kpc $< R <$ 18.5 kpc and 2.2–$2.8 \times 10^{10} M_\odot$), which may represent the remains of a cannibalized dwarf galaxy[1]. The most serious uncertainty in using gas as a tracer of the gravitational field arises from the need to adopt a model to derive the gas distribution. It is therefore highly desirable that the gravitational field in the outer Galaxy be investigated using young stars for which good distance

[1]Astrophysics, Cosmology and Gravity Centre, Astronomy Department, University of Cape Town, Rondebosch 7701, South Africa. [2]South African Astronomical Observatory, PO Box 9, Observatory 7935, South Africa. [3]Department of Astronomy, School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan.

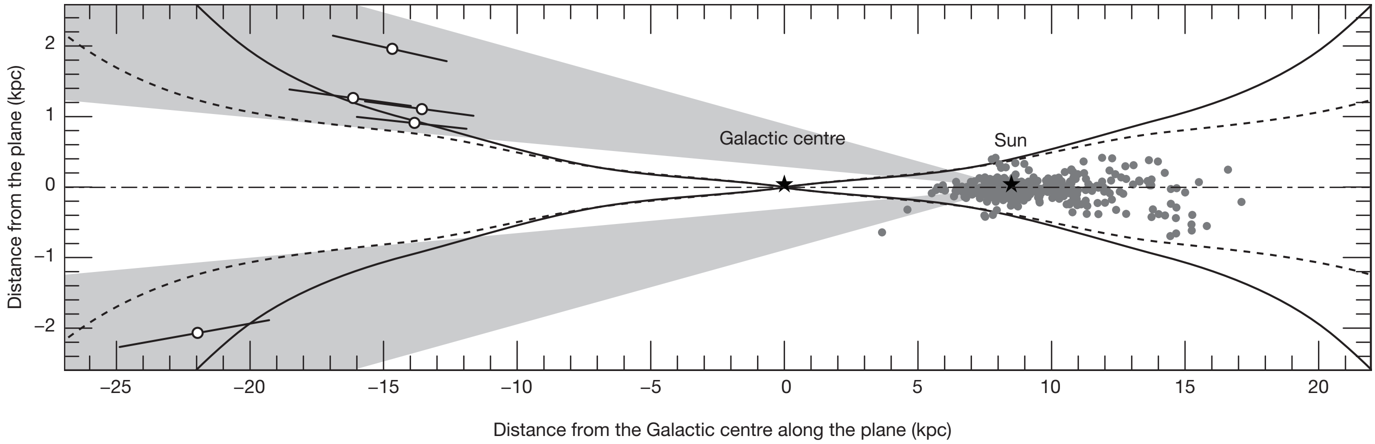


Figure 1 | Schematic of the Galaxy. The positions of the Cepheids (open circles with assumed maximum uncertainties of ±0.2 mag) are compared to the location of the H I gas. The solid and dashed curves are model fits, S and N1, respectively, from ref. 1 at three times the HWHM above and below the Galactic plane. We note that figures 1 and 2 of ref. 2 show the H I flare in the relevant region extending up to about 2 kpc. The dark grey points are previously known Galactic Cepheids[10] and the approximate regions surveyed by OGLE ($2 < |b| < 6$) are shown in light grey on either side of the plane. The positions of the Sun and Galactic centre are indicated by the star symbols.

estimates can be made. Classical Cepheid variables are by far the best stars for this purpose.

Studies of diffuse groups of B stars[12], which are even younger than Cepheids, are also consistent with a Galactic disk extending 15 kpc and 20 kpc from the centre, at Galactic latitude $b = -4^\circ$ and -7°, respectively. These stars are in the third Galactic quadrant near the place where the warp forces the Galactic plane to its greatest negative displacement from $b = 0^\circ$. So although these young stars are displaced from $b = 0^\circ$, they are in the local Galactic plane, and therefore tell us nothing about a flare.

The collection of stars now known as the 'Monoceros ring' has been interpreted as evidence for a warped disk[13], or alternatively as the remnant of a dwarf galaxy cannibalized by the Milky Way[14]. It is perhaps curious that the Cepheids discussed above are at the distance from the Galactic centre that one would expect the Monoceros ring to be, if indeed it were a complete circular ring around the Galaxy. The stellar population that makes up this so-called ring is generally considered to be old (>1 Gyr) and therefore different from the Cepheids (although there have been suggestions of an association with spiral arms[11]). Models[15] indicate that the ages of the youngest Cepheids discussed here are less than 130 Myr. The disputed origin of the Monoceros ring[16] is beyond the scope of this Letter. Nevertheless, we note that simulations that suggest that the ring is a consequence of the interaction of the Sgr dwarf galaxy with the Milky Way[17] do not predict any significant density of stars in the ring at the distance of the Cepheids under discussion.

Clearly, these Cepheids are just the tip of the iceberg. Further work on these stars and other 'standard candles' in the outer Galaxy will present new opportunities to probe the gravitational field and therefore the distribution of dark matter in the outer parts of our Galaxy.

METHODS SUMMARY

The Fourier coefficients listed for each light curve of the candidate Cepheids[3] were compared with those of classical Cepheids in the Large Magellanic Cloud (LMC)[18] to show that five of the stars with periods greater than one day fall clearly into the classical Cepheid class; we can therefore derive their distances from their luminosities.

The distances and the interstellar absorptions were derived together using pairs of colours (V and I or J and K_S). The results from the infrared magnitudes were adopted because the uncertainty due to interstellar reddening is significantly higher at shorter wavelengths. The detailed analysis indicates that the reddening law towards the Galactic centre is abnormal, as is well known[19]. Various sources of uncertainty on the distance moduli are discussed in detail in the Methods, but the reddening law and the exact values of the reddening are the primary contributors, which lead to our estimate of the upper limit to the uncertainty of ±0.2 magnitudes (mag). Radial velocities were determined by cross-correlation with a synthetic spectrum and the zero point of the velocity scale was confirmed by observation of two stars with known velocities. An approximate calculation can be made of the numbers of Cepheids expected by extrapolating from the solar neighbourhood and assuming a scale length of 3 kpc within the plane. If the Cepheids in the flared disk have the same scale height as the gas (577 pc) then we would expect about 18 to exist above a height of 1 kpc from the plane in the direction surveyed by OGLE. Given the uncertainties, this is consistent with the five Cepheids that we do find, particularly as we do not expect our sample to be complete.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper

Received 6 December 2013; accepted 11 March 2014.

1. Kalberla, P. M. W., Dedes, L., Kerp, J. & Haud, U. Dark matter in the Milky Way. II. The H I gas distribution as a tracer of the gravitational potential. *Astron. Astrophys.* **469,** 511–527 (2007).
2. Levine, E. S., Blitz, L. & Heiles, C. The vertical structure of the outer Milky Way H I disk. *Astrophys. J.* **643,** 881–896 (2006).
3. Soszyński, I. *et al.* The optical gravitational lensing experiment. The OGLE-III catalog of variable stars. XIV. Classical and type II Cepheids in the Galactic bulge. *Acta Astronaut.* **61,** 285–301 (2011).
4. Feast, M. W. in *Planets, Stars and Stellar Systems* (eds Oswalt, T. D. & Gilmore, G.) Vol. 5, 829 (Springer, 2013).
5. Dame, T. M. & Thaddeus, P. Molecular spiral arm in the far outer Galaxy. *Astrophys. J.* **734,** L24 (2011).
6. Ibata, R. A., Wyse, R. F. G., Gilmore, G., Irwin, M. J. & Suntzeff, N. B. The kinematics, orbit, and survival of the Sagittarius dwarf spheroidal galaxy. *Astron. J.* **113,** 634–655 (1997).
7. Frinchaboy, P. M. *et al.* A 2MASS all-sky view of the Sagittarius dwarf galaxy. VII. Kinematics of the main body of the Sagittarius dSph. *Astrophys. J.* **756,** 74 (2012).
8. Rich, R. M. in *Planets, Stars and Stellar Systems* (eds Oswalt, T. D. & Gilmore, G.) Vol. 5 271 (Springer, 2013).
9. Matsunaga, N. *et al.* Three classical Cepheid variable stars in the nuclear bulge of the Milky Way. *Nature* **477,** 188–190 (2011).
10. Fernie, J. D., Evans, N. R., Beattie, B. & Seager, S. A database of Galactic classical Cepheids. *Inform. Bull. Variable Stars* **4148** (1995).
11. Olling, R. P. On the usage of flaring gas layers to determine the shape of dark matter halos. *Astron. J.* **110,** 591–612 (1995).
12. Carraro, G., Vázquez, R. A., Costa, E., Perren, G. & Moitinho, A. The edge of the young Galactic disk. *Astrophys. J.* **718,** 683–694 (2010).
13. Ibata, R. A., Irwin, M. J., Lewis, G. F., Ferguson, A. M. N. & Tanvir, N. One ring to encompass them all: a giant stellar structure that surrounds the Galaxy. *Mon. Not. R. Astron. Soc.* **340,** L21–L27 (2003).
14. Yanny, B. *et al.* A low-latitude halo stream around the Milky Way. *Astrophys. J.* **588,** 824–841 (2003).
15. Bono, G. *et al.* Classical Cepheid pulsation models. X. The period-age relation. *Astrophys. J.* **621,** 966–977 (2005).
16. López-Corredoira, M. *et al.* Comments on the "Monoceros" affair http://arxiv.org/abs/1207.2749 (2012).
17. Purcell, C. W., Bullock, J. S., Tollerud, E. J., Rocha, M. & Chakrabarti, S. The Sagittarius impact as an architect of spirality and outer rings in the Milky Way. *Nature* **477,** 301–303 (2011).
18. Soszyński, I. *et al.* The optical gravitational lensing experiment. The OGLE-III catalog of variable stars. II. Type II Cepheids and anomalous Cepheids in the Large Magellanic Cloud. *Acta Astronaut.* **58,** 293–312 (2008).
19. Nishiyama, S. *et al.* Interstellar extinction law in the J, H, and K_s bands toward the Galactic center. *Astrophys. J.* **638,** 839–846 (2006).

Acknowledgements The observations reported in this Letter were obtained with the Japanese–South African IRSF at Sutherland and the SALT. M.W.F., J.W.M. and P.A.W. gratefully acknowledge research grants from the South African National Research Foundation.

Author Contributions M.W.F. coordinated the project and conducted the analysis. J.W.M. reduced the spectroscopic observations from SALT and determined the radial velocities. N.M. made and analysed the IRSF observations. All four authors contributed to the explanation and the discussion.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.W.F. (mwf@ast.uct.ac.za).

METHODS

In the following, we describe how the stars were identified as classical Cepheids by comparison with similar stars in the LMC[18]. We then go on to derive distances, taking into account the well known abnormal reddening law towards the Galactic centre[19].

Identifying classical Cepheids. A problem when studying Cepheids is that it is not always easy from available photometry to distinguish classical Cepheids (type I) from other objects, for example, anomalous or type II Cepheids (BL Her stars, W Vir stars). This issue has been discussed in the context of distant Cepheids towards the anti-centre[20] (that is, the direction in the Galaxy that is opposite from the centre, viewed from our perspective). In the interior region of the Galaxy, and particularly in the direction of the bulge, this is likely to be a significant problem. Fortunately it is possible to distinguish between some classes of stars using the Fourier coefficients of their light curves, and these are listed for the OGLE[3] Cepheids towards the bulge.

The main Fourier parameters for the *I*-band light curves of the five stars discussed here (Extended Data Table 1) can be compared with plots of the Fourier coefficient ratios R_{21}, R_{31} and phase differences φ_{21}, φ_{31} (where the subscripts denote the order of the cosine curve fit) against period for various classes of variable star in the LMC[18] and this enables us to classify these five securely as classical Cepheids. Other possible Cepheids in the OGLE bulge catalogue have characteristics that suggest they belong to the anomalous Cepheid class, are possible type II Cepheids or else their classification is doubtful.

Photometry. The infrared photometry (Extended Data Table 2) was carried out using the 1.4-m IRSF and the SIRIUS camera at Sutherland[21]. Each of the targets was observed once on 2012 May 6 (Universal time, UT) with an exposure time of 25 s (5 s times five exposures at dithered positions). The photometry was extracted using the Image Reduction and Analysis Facility (IRAF) package DAOPHOT (http://iraf.noao.edu) and standardized by comparison with nearby stars from the 2MASS point source catalogue[22]. The uncertainties for the brightest and faintest of the Cepheids range from 0.02–0.07 mag at *J*, 0.02–0.03 mag at *H* and 0.02–0.04 mag at K_S, respectively. These are significantly less than the uncertainties on the 2MASS measures, where they exist, for the same sources. We use these single-epoch *J*, *H* and K_S measurements to estimate the distance, noting that the near-infrared amplitudes of these short-period stars will be small (<0.1 mag; ref. 23).

Distances and interstellar absorptions. In general there are severe problems in dealing with observations of distant stars in the Galactic plane close to or beyond the centre because of the large and uncertain amounts of interstellar extinction in these directions. Cepheids offer an important advantage in this regard in that distances can be derived from relations that allow the reddening and the distance to be determined together and unambiguously when observations in two colours are available—for example, *V* and *I* or *J* and K_S—provided the reddening law is known. Recent work[19] has indicated that the law of reddening is different towards the Galactic bulge from that adopted elsewhere[24] and here we use:

$$A_{\mathrm{K}} = (0.494 \pm 0.006)E_{(J-K)}$$

from ref. 19 and

$$A_{\mathrm{I}} = (1.125 \pm 0.09)E_{(V-I)}$$

which was found by the same method[25]. It should be noted that the relation in *V* and *I* may be somewhat more complex than the one given[25].

Adopting period–luminosity relations in *V* and *I* (as derived[26] by the OGLE group) and *J* and K_S (derived[27] for Cepheids with $0.4 < \log P < 1.0$) from the LMC together with an LMC distance modulus of 18.5 mag and interstellar extinction values[28] of $A_V = 0.22$ mag, $A_I = 0.13$ mag, $A_J = 0.06$ mag and $A_K = 0.02$ mag for the LMC direction, we then have the distance modulus μ_0 for a Cepheid with a pulsation period *P*:

$$\mu_0 + A_V = V + 2.762\log P + 1.190 \tag{1}$$

$$\mu_0 + A_I = I + 2.959\log P + 1.751 \tag{2}$$

and

$$\mu_0 + A_J = J + 3.138\log P + 2.109 \tag{3}$$

$$\mu_0 + A_K = K_S + 3.284\log P + 2.383 \tag{4}$$

Combining these pairs of equations with the reddening law given above leads to the two estimates of the distance modulus, $(\mu_0)_{VI}$ and $(\mu_0)_{JK}$—derived, respectively, from equations (1) and (2) and equations (3) and (4)—and the interstellar extinction corrections, A_I and A_K (Extended Data Table 2).

Uncertainties in the distances. The LMC period–luminosity relations that we used are well defined[4]. Their absolute calibration is based on the LMC distance modulus, which has been determined in a number of ways. The uncertainty in the adopted value (about 0.04 mag or 2%) is negligible for our discussion. The mean OGLE *VI* magnitudes derived from their extensive observations have negligible error. Because of the small pulsation amplitudes of the Cepheids in the infrared region of the spectrum, the error on our *J* and K_S magnitudes is ≤0.05 mag.

Possible metallicity effects on the period–luminosity relations have been much discussed[4]. Nothing is known about the metallicities of Cepheids behind the Galactic centre, but those of Cepheids in the outer disk of the Galaxy in the general direction of the anti-centre[29], and at comparable distances from the centre to those discussed here, have a mean logarithmic iron-to-hydrogen ratio [Fe/H] = −0.60 ± 0.12, which is intermediate between those of the LMC and the Small Magellanic Cloud[30].

The difference in distance between the Small Magellanic Cloud and the LMC derived from *J* and K_S observations of Cepheids[31] agrees with values measured in other ways, without the application of metallicity corrections. Furthermore, Hubble Space Telescope parallaxes of Galactic Cepheids[32] (with [Fe/H] ≈ 0) agree with the LMC modulus adopted without the application of any metallicity corrections (they give 18.52 ± 0.03 mag from *V* and *I* and 18.47 ± 0.03 mag from K_S). The various factors indicate that any residual metallicity effects on the distances derived for these Cepheids will be very small.

A potential source of uncertainty is in the width of the period–luminosity relations. This width is due to the fact that, at a given period, a Cepheid brighter than the average is also bluer. This leads to the smaller-than-average *V* being compensated by a lower-than-average derived apparent absorption. It is clear that the uncertainty in the modulus due to the spread in colour at a given period is[4]:

$$\sigma(\mu_0) = (\beta_1 - \beta_2)\sigma_{(V-I)0} \tag{5}$$

where β_1 is the colour coefficient of a (nearly dispersionless) period–luminosity–colour relation in (*V*, *I*) and β_2 is the ratio of total to selective absorption. For the Cardelli[24] law of reddening, which is often used, $\beta_1 \approx \beta_2$. Thus, any uncertainty due to the width of the period–luminosity relation in our case comes from the change in β_2 for the bulge, which is 0.33. The scatter in $V - I$ at a given period[33] is 0.08, which would result in $\sigma(\mu_0) = 0.03$ in equation (5). In the infrared, the widths of the period–luminosity relations are lower and will not introduce significant uncertainty.

Interstellar reddening is a source of error and, as pointed out above, the evidence points to an abnormal reddening law in the direction of these stars. The uncertainty in this reddening law in JK_S is small; this, together with the low extinction in the infrared, leads to only a small uncertainty in the distance modulus (0.003 mag for the most heavily reddened star). Even if, contrary to the evidence, we used the Cardelli[24] law of reddening, the change in distance moduli would not affect our conclusions. In that case, the modulus of our most reddened star would decrease by 0.28 mag (a change of distance from 24.4 kpc to 21.4 kpc) and the moduli of the other stars would decrease by an average of 0.12 mag (1.25 kpc). Owing to the greater absorption in *V* and *I* and the greater uncertainty in the reddening law, the uncertainties in the derived distances are greater. The uncertainty in the reddening coefficient leads to an uncertainty of 0.25 mag in the modulus of the most heavily reddened star and a mean of 0.12 mag in the other cases. If, contrary to expectations, a Cardelli reddening law had been adopted, the modulus of the most heavily absorbed star would have decreased by 0.90 mag and those of the others by a mean of 0.42 mag. Clearly, reddening uncertainties in *VI* are much more important than in *JK*.

Summary of adopted distances and their uncertainties. In the main paper we adopt the distances derived from the *J* and K_S magnitudes (see Table 1), because they are the more accurate values. The above discussion indicates that the errors in those distance moduli are: 0.04 mag from the absolute calibration; ≤0.05 mag due to the pulsation amplitude; and negligible amounts from the period–luminosity relation width, metallicity effects and uncertainties in the Nishiyama reddening law. If the Cardelli reddening law were applied to these stars their moduli would be reduced by a mean of 0.15 mag, but a change this large seems to be ruled out by observations. The systematic uncertainties overwhelm the rather small statistical errors, so we do not attempt to assign individual errors to distances. We consider 0.2 mag to be a very conservative estimate of the total error of an individual modulus (random plus systematic) and this is what is illustrated in Fig. 1, but we fully expect the errors to be less than this. In the case of the moduli from *V* and *I*, the main uncertainty is from the coefficient of the reddening law and complications in deriving this have been noted[25]. We simply mention here that with the adopted law the *VI* moduli are 0.20 mag larger than the JK_S values adopted, whereas with a Cardelli law they are 0.31 mag smaller, suggesting that a less extreme variation from the Cardelli law applies to these stars.

Radial velocities. Our spectra (Extended Data Table 3) were obtained with the Robert Stobie spectrograph on the SALT. A volume phase holographic grating of 1,300 lines mm^{-1} was used to cover the wavelength range 7,800–9,600 Å, putting the Ca II triplet on the middle charge-coupled device of the detector. The resolution is 3.4 Å with a projected slit width of 1.5 arcsec.

Radial velocities were obtained by cross-correlation of the spectra with a synthetic spectrum taken from the library assembled for the RAVE experiment[34]. Two

stars with known radial velocities[35] were used as a check on the radial velocity zero point. The measured velocities for these two stars are 34.7 km s^{-1} and -12.0 km s^{-1}, respectively. Mean radial velocity errors due to photon statistics are 10 km s^{-1}, so the radial velocity zero point seems to be secure. The measured heliocentric velocities have been corrected for stellar pulsation adopting a standard velocity curve for short-period Cepheids with a semi-amplitude of 20 km s^{-1} (for example, figure 6 of ref. 36). The mean heliocentric velocity after correction is 4 ± 8 km s^{-1} compared with 13 ± 7 km s^{-1} before correction. This indicates that uncertainties in the correction will not affect our conclusion regarding the mean radial outward velocity of this group of stars and that the error given in the main text is realistic.

Galactic structure. In the main paper and in the following we adopt a distance from the Sun to the Galactic centre of 8.5 kpc and a flat rotation curve with a velocity $\theta = 220$ km s^{-1}, to allow for a direct comparison with models describing the H I gas behaviour in the outer Galactic disk[2]. Plausible changes[37] in these values will not affect our conclusions.

The heliocentric distances of the Cepheids (D values in Table 1) are comparable with that of the Sgr dwarf galaxy (about 24 kpc), and a tidal stream from this system crosses the Galactic plane, behind the Galactic centre, close to the Galactic bulge at positive Galactic longitude. RR Lyrae variables belonging to this stream have recently been found in our field[38] at a distance of around 27 kpc. The possibility that the Sgr system contains stars as young as about 100 Myr has also been raised[39]. This is the expected age of short-period classical Cepheids[15], so we cannot rule out the possibility that our stars belong to the Sgr system on the basis of photometry alone, and kinematic information is essential. Because the heliocentric radial velocities of Sgr dwarf members are about 150 km s^{-1} (refs 6 and 7), it is clear from the velocities in Table 1 that our Cepheids belong instead to the far outer parts of the Galactic disk.

The possible association of the Cepheids with the far outer molecular spiral arm[5] was raised in the main paper. This arm lies at positive Galactic longitudes (in the first quadrant). At $l = 13^\circ.25$, the lowest longitude at which the carbon monoxide (CO) was measured, the estimated distance[5] is $D = 23$ kpc, corresponding to $R = 14.5$ kpc; the exact distances are sensitive to the kinematic model. These are somewhat less than the D and R values in Table 1. Our values (in Table 1) of course refer to a region where there is no information from the gas. Adopting an alternative kinematic model[2] with elliptical orbits will lead to larger derived distances of the gas.

The five Cepheids are concentrated in a relatively small part of the area covered by the OGLE survey at positive longitudes. It is possible that variable interstellar absorption over the field could account for this. However, it seems more likely that it is due to real clumping, such as is common for young objects in spiral arms. The far outer molecular arm has not yet been seen emerging from the Galactic centre region at negative longitudes.

We note that our stars are at Galactocentric distances comparable to, but greater than, those of a small number of masers defining an outer arm in the general direction of the anti-centre[40]. The Cepheids whose radial velocities were studied in the general direction of the anti-centre are also at somewhat shorter distances (mean $R = 12.9$ kpc)[20].

The Sun–Cepheid–Galactic centre angle is small for all of our stars (0° to 2.7°, as measured in the Galactic plane). Thus, the corrected radial velocity ρ primarily measures motion that is radial from the Galactic centre and the five Cepheids give a mean $\rho = 23 \pm 9$ km s^{-1}, indicating a significant mean outward radial motion. This would not be inconsistent with a value of 9 km s^{-1} predicted in the model[2]. It should also be noted that in the general solar neighbourhood systematic deviations from circular motion of about 10 km s^{-1} are known for OB stars and Cepheids in regions of around a kiloparsec radius[41]. Therefore, our result does not necessarily imply any significant general deviation from circular orbits. The uncertainty in this conclusion is related to the small number of objects involved rather than to the uncertainties in estimating mean Cepheid velocities.

In the general anti-centre region at somewhat smaller Galactocentric distances no evidence was found[20] for a general outward velocity, though curiously the three Cepheids in that study[20] with Galactic latitudes of $|\Delta l| < 10^\circ$ from the anti-centre have a mean positive radial velocity of 10 ± 4 km s^{-1}. Despite the small number of our stars the result would be in conflict with an outward motion of the local standard of rest of about 14 km s^{-1} which has sometimes been suggested[42] to explain the Galactic asymmetry of the H I velocities.

Number of Cepheids observed and expected. With such a small number of Cepheids in our sample it is impossible to carry out a detailed study of the space distribution at their distance from the Galactic centre. However, the following rough calculation shows that their presence far from the Galactic plane requires the presence of a flared disk.

Consideration of the number of Cepheids in the solar neighbourhood suggests that the expected number of such stars in a column perpendicular to the Galactic plane and with a cross-sectional area of one square kiloparsec is about 60. With a scale height of 86 pc, as for the gas[2] (HWHM 60 pc) and taking into account that the area on the Galactic plane, between $D = 20$ kpc and $D = 30$ kpc, covered by the OGLE survey of (slightly less than) 44 kpc^2, the number of Cepheids expected in that survey with $z > 1$ kpc is approximately 0.008. This is for solar neighbourhood densities. With a disk scale length of 3 kpc, the drop in density from 8.5 kpc to 15 kpc is a factor of 9 and the expected number of Cepheids is about 0.001 (that is, for an unflared disk Cepheids are not expected). However, at the distance of our Cepheids the scale height of the gas is about 577 pc (HWHM 400 pc) and if the Cepheids follow the gas we predict the existence of about 18 in the relevant region. This calculation is obviously quite uncertain, but it is sufficient to show that our conclusion that these Cepheids are in the outer regions of a flared disk of scale height similar to that of the gas is entirely plausible. We see no other satisfactory explanation for these stars. We cannot rule out the possibility that a few more of the OGLE variables are classical Cepheids.

Owing to the small numbers, the likely effects of non-uniform interstellar absorption and the fact that young objects are expected to be found in groups rather than uniformly distributed over the field, it is not feasible to draw any strong conclusion from the fact that these Cepheids are confined to the positive longitude side of the OGLE field or that four of the five stars are at northern latitudes, despite the fewer OGLE fields there.

20. Pont, F., Queloz, D., Bratschi, P. & Mayor, M. Rotation of the outer disc from classical cepheids. *Astron. Astrophys.* **318,** 416–428 (1997).
21. Nagayama, T. *et al.* SIRIUS: a near infrared simultaneous three-band camera. *Proc. SPIE* **4841,** 459–464 (2003).
22. Skrutskie, M. F. The Two Micron All Sky Survey (2MASS). *Astron. J.* **131,** 1163–1183 (2006).
23. Laney, C. D. & Stobie, R. S. Light-curve systematics of Cepheids in the infrared. *Mon. Not. R. Astron. Soc.* **260,** 408–414 (1993).
24. Cardelli, J. A., Clayton, G. C. & Mathis, J. S. The relationship between infrared, optical, and ultraviolet extinction. *Astrophys. J.* **345,** 245–256 (1989).
25. Nataf, D. M. *et al.* Reddening and extinction toward the Galactic Bulge from OGLE-III: The inner Milky Way's $R_V \sim 2.5$ extinction curve. *Astrophys. J.* **769,** 88 (2013).
26. Soszyński, I. *et al.* The Optical Gravitational Lensing Experiment. The OGLE-III catalog of variable stars. II. Type II Cepheids and anomalous Cepheids in the Large Magellanic Cloud. *Acta Astronaut.* **58,** 163–185 (2008).
27. Matsunaga, N., Feast, M. W. & Soszyński, I. Period-luminosity relations of type II Cepheids in the Magellanic clouds. *Mon. Not. R. Astron. Soc.* **413,** 223–234 (2011).
28. Caldwell, J. A. R. & Coulson, I. M. BVI reddenings of Magellanic cloud Cepheids. *Mon. Not. R. Astron. Soc.* **212,** 879–888 (1985).
29. Yong, D., Carney, B. W., Teixera de Almeida, M. L. & Pohl, B. L. Elemental abundance ratios in stars of the outer Galactic disk. III. Cepheids. *Astron. J.* **131,** 2256–2273 (2006).
30. Groenewegen, M. A. T. Baade-Wesselink distances to Galactic and Magellanic cloud Cepheids and the effect of metallicity. *Astron. Astrophys.* **550,** A70 (2013).
31. Inno, L. *et al.* On the distance of the Magellanic clouds using Cepheid NIR and optical-NIR period-Wesenheit relations. *Astrophys. J.* **764,** 84 (2013).
32. Benedict, G. F. *et al.* Hubble Space Telescope fine guidance sensor parallaxes of Galactic Cepheid variable stars: period-luminosity relations. *Astron. J.* **133,** 1810–1827 (2007).
33. Sandage, A., Tammann, G. A. & Reindl, B. New period-luminosity and period-colour relations of classical Cepheids. II. Cepheids in the LMC. *Astron. Astrophys.* **424,** 43–71 (2004).
34. Zwitter, T., Castelli, F. & Munari, U. An extensive library of spectra covering the far red, RAVE and GAIA wavelength ranges. *Astron. Astrophys.* **417,** 1055–1062 (2004).
35. Johnson, C. I., Rich, R. M., Fullbright, J. P., Valenti, E. & McWilliam, A. Alpha enhancement and the metallicity distribution function in Plaut's window. *Astrophys. J.* **732,** 108 (2011).
36. Marconi, M., Molinaro, R., Ripepi, V., Musella, I. & Brocato, E. Theoretical fit of Cepheid light and radial velocity curves in the Large Magellanic Cloud cluster NGC 1866. *Mon. Not. R. Astron. Soc.* **428,** 2185–2197 (2013).
37. Sofue, Y. in *Planets, Stars and Stellar Systems* **5** (eds Oswalt, T. D. & Gilmore, G.) 985 (Springer, 2013).
38. Jin, S., Grebel, E. K. & Haschke, R. Seeing Sagittarius through the bulge: tracing stellar stream stars in OGLE. *IAU Symp.* **298,** 415 (2013).
39. Siegel, M. H. *et al.* The ACS survey of Galactic globular clusters: M54 and young populations in the Sagittarius dwarf spheroidal galaxy. *Astrophys. J.* **667,** L57–L60 (2007).
40. Brunthaler, A. *et al.* The Bar and Spiral Structure Legacy (BeSSeL) survey: mapping the Milky Way with VLBI astrometry. *Astron. Nachr.* **332,** 461–466 (2011).
41. Feast, M. W. The radial velocities of distant Cepheids and galactic models. *Mon. Not. R. Astron. Soc.* **136,** 141–164 (1967).
42. Kerr, F. J. & Westerhout, G. in *Galactic Structure* (eds Blaauw, A. & Schmidt, M.) 167 (Univ. Chicago Press, 1965).

Extended Data Table 1 | Fourier coefficients for the *I*-band light curves of the Cepheids

OGLE #	Log P	R_{21}	φ_{21}	R_{31}	φ_{31}
01	0.414	0.488	4.458	0.302	2.807
02	0.307	0.522	4.263	0.354	2.444
03	0.092	0.099	3.603	0.157	1.706
05	0.579	0.443	4.753	0.212	3.242
32	0.572	0.436	4.556	0.232	2.791

The amplitude (A) ratios $R_{n1} = A_n/A_1$ and phase differences $\varphi_{n1} = \varphi_n - n\varphi_1$ are listed, where A_n and φ_n are parameters of the truncated Fourier series fitted to the photometric data[3,18]. The subscripts refer to the order of the fit, so that $n = 1$ is the fundamental, $n = 2$ is the first harmonic and so on. P is the pulsation period in days.

Extended Data Table 2 | Photometry of the Cepheids

OGLE #	01	02	03	05	32
V	20.800	17.482	18.207	17.675	16.731
I	17.382	15.682	16.390	15.374	15.047
J	15.28	14.34	15.25	13.67	14.04
H	14.03	13.79	14.61	13.03	13.42
K_S	13.79	13.63	14.34	12.74	13.33
JD-2456053	0.61250	0.61389	0.61458	0.61667	0.64514
$(\mu_0)_{VI}$	17.24	17.01	17.02	17.01	17.35
$(\mu_0)_{JK}$	16.96	16.83	16.72	16.74	17.42
A_I	3.12	1.33	1.39	1.83	1.14
A_K	0.57	0.19	0.31	0.28	0.17

For each star the OGLE mean V and I magnitudes and the single epoch IRSF J, H and K_S magnitudes for observations made on the given Julian date (JD) are listed. $(\mu_0)_{VI}$ is the reddening-corrected distance modulus calculated from equations (1) and (2) and $(\mu_0)_{JK}$ is the reddening-corrected distance modulus calculated from equations (3) and (4). A_I and A_K are the interstellar extinction values at I and K_S respectively, calculated simultaneously with the distance moduli.

Extended Data Table 3 | Journal of spectroscopic observations

Object	HJD – 2450000.0	Phase/(V_R)
OGLE-BLG-CEP-01	6433.40318	0.906
OGLE-BLG-CEP-02	6498.44740	0.914
OGLE-BLG-CEP-03	6463.56319	0.545
OGLE-BLG-CEP-05	6409.47279	0.845
OGLE-BLG-CEP-32	6498.47404	0.727
2MASS J18181710-3401088	6498.48681	(+35 kms^{-1})
2MASS J18182553-3349465	6498.50206	(+0 kms^{-1})

For each object named, the heliocentric Julian date (HJD) when the spectrum was obtained and the phase of the Cepheid variations is listed. For the two reference stars, the catalogue radial velocities (V_R) are given.

LETTER

doi:10.1038/nature13252

Tracking excited-state charge and spin dynamics in iron coordination complexes

Wenkai Zhang[1], Roberto Alonso-Mori[2], Uwe Bergmann[2], Christian Bressler[3], Matthieu Chollet[2], Andreas Galler[3], Wojciech Gawelda[3], Ryan G. Hadt[4], Robert W. Hartsock[1,4], Thomas Kroll[4], Kasper S. Kjær[5,6], Katharina Kubiček[7,8], Henrik T. Lemke[2], Huiyang W. Liang[1,4], Drew A. Meyer[1,4], Martin M. Nielsen[6], Carola Purser[1], Joseph S. Robinson[2], Edward I. Solomon[4,9], Zheng Sun[1], Dimosthenis Sokaras[9], Tim B. van Driel[6], György Vankó[10], Tsu-Chien Weng[9], Diling Zhu[2] & Kelly J. Gaffney[1]

Crucial to many light-driven processes in transition metal complexes is the absorption and dissipation of energy by 3*d* electrons[1–4]. But a detailed understanding of such non-equilibrium excited-state dynamics and their interplay with structural changes is challenging: a multitude of excited states and possible transitions result in phenomena too complex to unravel when faced with the indirect sensitivity of optical spectroscopy to spin dynamics[5] and the flux limitations of ultrafast X-ray sources[6,7]. Such a situation exists for archetypal polypyridyl iron complexes, such as $[Fe(2,2'\text{-bipyridine})_3]^{2+}$, where the excited-state charge and spin dynamics involved in the transition from a low- to a high-spin state (spin crossover) have long been a source of interest and controversy[6–15]. Here we demonstrate that femtosecond resolution X-ray fluorescence spectroscopy, with its sensitivity to spin state, can elucidate the spin crossover dynamics of $[Fe(2,2'\text{-bipyridine})_3]^{2+}$ on photoinduced metal-to-ligand charge transfer excitation. We are able to track the charge and spin dynamics, and establish the critical role of intermediate spin states in the crossover mechanism. We anticipate that these capabilities will make our method a valuable tool for mapping in unprecedented detail the fundamental electronic excited-state dynamics that underpin many useful light-triggered molecular phenomena involving 3*d* transition metal complexes.

The femtosecond duration of the intense hard X-ray pulses generated by the LCLS (Linac Coherent Light Source) X-ray free-electron laser[16,17] creates the opportunity to study spin dynamics with iron 3*p*–1*s* (Kβ) X-ray fluorescence spectroscopy[18,19]. Figure 1 shows diagrams of the measurement technique and relevant energy levels (Fig. 1–c), a 'ball-and-stick' representation of the $[Fe(2,2'\text{-bipyridine})_3]^{2+}$ complex (Fig. 1d), and the dependence of photoexcited spin crossover dynamics on the Fe–ligand distance (Fig. 1e). Given the roughly 100 femtosecond (fs) time resolution of the measurement[17], the subfemtosecond lifetime of the iron 1*s* core hole makes X-ray fluorescence an effectively instantaneous probe[20]. A variety of distinct electronic excited states, including singlet and triplet metal-to-ligand charge transfer states (^{1,3}MLCT), triplet ligand field excited states (^{3}T) and quintet ligand field excited states ($^5T_2$) have been proposed to participate in the spin crossover mechanism[6,8,10,11,21,22] (Fig. 1e). Distinguishing electronic excited states with different charge and spin density, such as the ^{1,3}MLCT, ^{3}T and $^5T_2$ states listed above, represents a critical step in characterizing the spin crossover mechanism.

Figure 2a shows the sensitivity of the iron Kβ fluorescence spectrum to the 3*d* spin moment, a sensitivity that results from the exchange interaction between the 3*p* and 3*d* electrons[18,19,23–25]. Equally important, the ground-state spectra of iron coordination complexes with different ligation, but the same iron spin moment, exhibit similar Kβ fluorescence spectra. This insensitivity of Kβ fluorescence spectroscopy to the details of the coordinating ligands and the local symmetry of the complex has previously been used to characterize the electronic ground-state spin moment of a variety of molecular systems[19,25]. We note that the insensitivity of the Kβ fluorescence spectrum to the electronic properties of the ligand means that the spectrum cannot be used to distinguish between singlet and triplet MLCT states. We utilize these spectra of distinct spin configurations to model transient difference spectra—that is, the time and energy dependence of the fluorescent amplitude difference between excited-state and ground-state spectra. Figure 2b shows the model complex difference spectra generated from the ground-state spectra of the relevant excited-state spin configurations and the singlet ground state. These model complex difference spectra confirm that each excited-state spin moment generates a distinct difference spectrum that cannot be reproduced by a linear combination of the other difference spectra (see Fig. 2, Extended Data Fig. 1 and Methods for details).

The time-resolved Kβ fluorescence spectra provide the sensitivity to spin dynamics needed to answer a critical question regarding the spin crossover mechanism: does the $^5T_2$ state form directly from the ^{1,3}MLCT state[6,13,26], or does spin crossover involve a ^{3}T transient[8,10]? Ultraviolet–visible transient absorption[13,14], time-resolved luminescence[13], and time-resolved iron K-edge XANES[6] have been used to characterize the spin crossover dynamics of $[Fe(2,2'\text{-bipyridine})_3]^{2+}$, and the similar rates measured for ^{3}MLCT decay and $^5T_2$ formation were attributed to the ^{3}MLCT excited state converting directly to the $^5T_2$ excited state, although a conversion including transient triplet states was also considered[6]. Potential energy surfaces calculated for this system allow either mechanism to proceed with minimal reaction barriers[21,22], but cannot explain why the ^{1,3}MLCT and $^5T_2$ states should be strongly coupled: the leading order spin–orbit interaction cannot couple the ^{1,3}MLCT and $^5T_2$ states because a transition between these states requires the excitation of two electrons on two distinct centres, whereas spin–orbit coupling is predominantly a single-centre, one-electron operator[22].

Figure 2c, d shows the transient difference spectra for $[Fe(2,2'\text{-bipyridine})_3]^{2+}$ measured for a 50-fs and a 1-ps (picosecond) time delay. The spectrum in Fig. 2d clearly demonstrates the ease of identifying the $^5T_2$ state with the Kβ fluorescence spectrum. Determining whether spin crossover from the ^{1,3}MLCT to the $^5T_2$ proceeds through a transient ^{3}T state proves more challenging because the relaxation dynamics do not lead to a time regime where the majority of the excited molecules reside in the ^{3}T excited state. The significant difference between the spectra in Fig. 2c and d, however, clearly demonstrates the presence of excited-state species other than the $^5T_2$ state. With statistically rigorous kinetic modelling, ^{1,3}MLCT, ^{3}T and $^5T_2$ states can be clearly distinguished in the relaxation dynamics probed with Kβ fluorescence.

The ability to spectroscopically distinguish between ^{1,3}MLCT, ^{3}T and $^5T_2$ electronic excited states allows the spin crossover mechanism

[1]PULSE Institute, SLAC National Accelerator Laboratory, Stanford University, Stanford, California 94305, USA. [2]LCLS, SLAC National Accelerator Laboratory, Menlo Park, California 94025, USA. [3]European XFEL, D-22761 Hamburg, Germany. [4]Department of Chemistry, Stanford University, Stanford, California 94305, USA. [5]Centre for Molecular Movies, Niels Bohr Institute, University of Copenhagen, DK-2100 Copenhagen, Denmark. [6]Centre for Molecular Movies, Department of Physics, Technical University of Denmark, DK-2800 Lyngby, Denmark. [7]Max Planck Institute for Biophysical Chemistry, 37077 Göttingen, Germany. [8]Deutsches Elektronen Synchrotron, Notkestraße 85, 22607 Hamburg, Germany. [9]SSRL, SLAC National Accelerator Laboratory, Menlo Park, California 94025, USA. [10]Wigner Research Centre for Physics, Hungarian Academy of Sciences, H-1525 Budapest, Hungary.

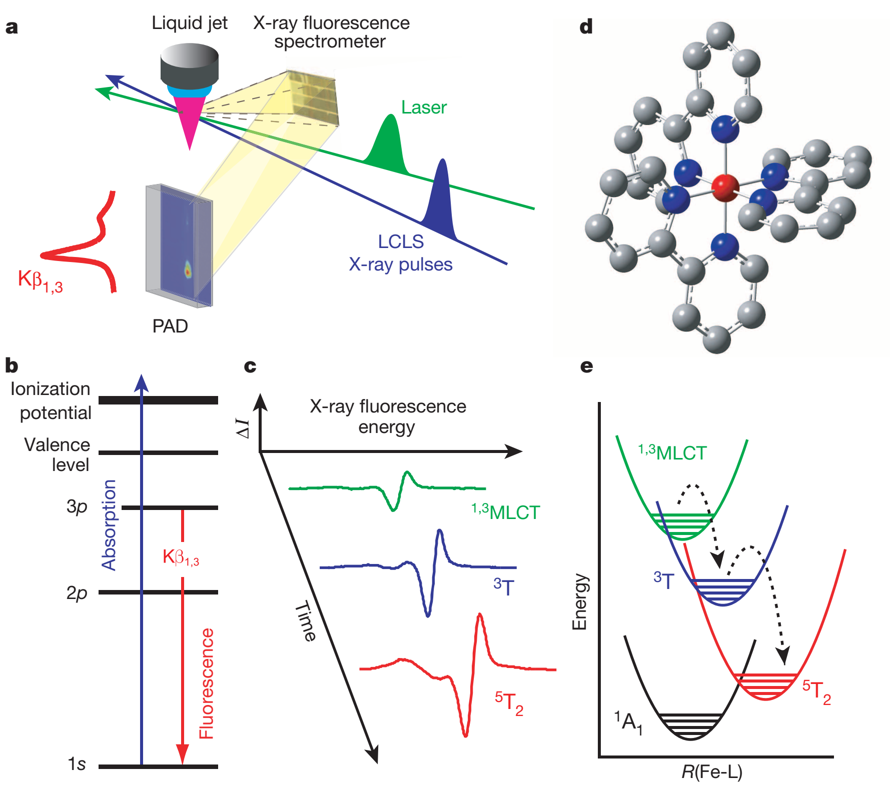


Figure 1 | Schematic depiction of ultrafast X-ray fluorescence detection of spin crossover dynamics. **a**, Experimental set-up involving liquid jet for sample replenishment, optical laser pump, and 8-keV X-ray beam for generating X-ray fluorescence measured with a dispersive crystal spectrometer. **b**, Energy level diagram for Kβ fluorescence involving photo-ionization of a 1*s* electron and X-ray fluorescence originating from the transition of a 3*p* electron to the 1*s* hole. **c**, Schematic diagram of how the spin crossover dynamics influence the time-dependent Kβ fluorescence difference spectra. **d**, Molecular structure of $[Fe(2,2'\text{-bipyridine})_3]^{2+}$ (red, Fe atom; blue, N; grey, C; H not shown). **e**, A schematic drawing of the potential energy surfaces involved in the spin crossover dynamics.

to be determined from the time evolution of the iron Kβ fluorescence spectrum. The time-resolved difference spectra, model fits of the difference spectra, and the parameters extracted from the fit can be found in Fig. 3, Extended Data Figs 2–4 and Extended Data Table 1. We have fitted the difference spectra to two distinct models: one where the ^{1,3}MLCT decays directly to a $^5T_2$ excited state and one where the ^{1,3}MLCT relaxes to the $^5T_2$ state via a ^3T transient. Figure 3b, c shows the time-dependent difference signal measured at two X-ray fluorescence energies: 7,061 eV, where the difference signal is largest, and 7,054 eV, where the triplet model complex has a spectral signature clearly distinct from the ^{1,3}MLCT and $^5T_2$ states as shown in Fig. 2b. The fits in Fig. 3b, c have been determined from a global analysis of the full time-dependent spectra. The statistical significance of the more complex kinetic model involving the triplet transient can be determined from an *F*-test comparison of the two models (described in Methods). The reduction in residuals achieved with the model containing the triplet transient is sufficient to reject the direct $^{1,3}\text{MLCT}\rightarrow{}^5T_2$ model with greater than 95% confidence. Note that the successful use of a kinetic model to describe subpicosecond

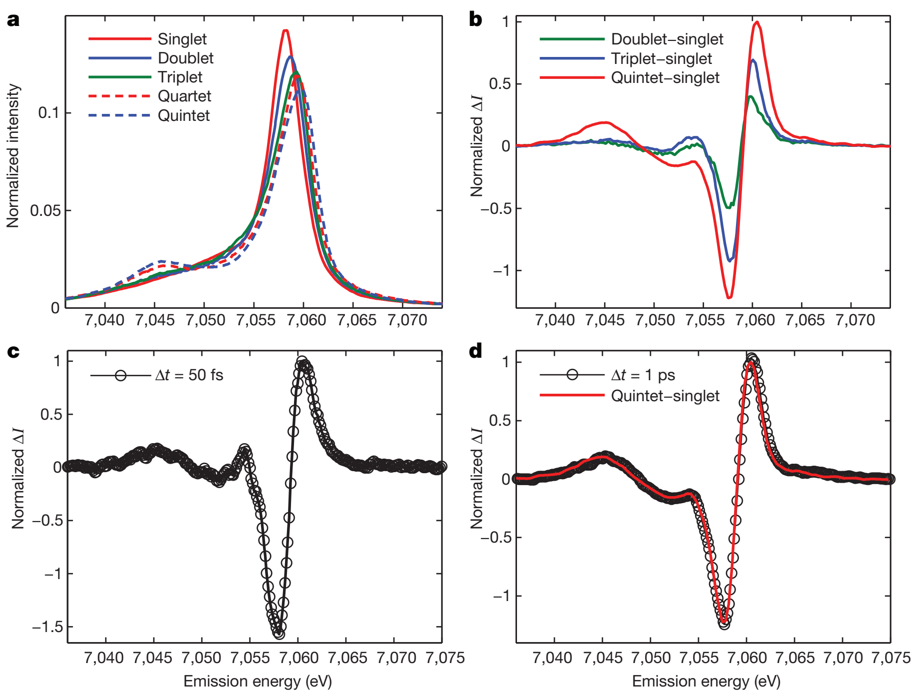


Figure 2 | Spin-dependent iron Kβ fluorescence spectra. **a**, The Kβ fluorescence spectra of ground-state iron complexes with different spin moments: singlet ($[Fe(2,2'\text{-bipyridine})_3]^{2+}$, red), doublet ($[Fe(2,2'\text{-bipyridine})_3]^{3+}$, blue), triplet (iron(II) phthalocyanine, green), quartet (iron(III) phthalocyanine chloride, red dashed), and quintet ($[Fe(\text{phenanthroline})_2(NCS)_2]$, blue dashed). **b**, Model complex difference spectra for the ^{1,3}MLCT, ^3T and $^5T_2$ excited states constructed by subtracting the singlet model complex spectrum from the doublet, triplet and quintet model complex spectra shown in **a**. **c**, Kβ transient difference spectra obtained at 50-fs time delay for $[Fe(2,2'\text{-bipyridine})_3]^{2+}$ (black circles). The best fit of this difference spectra can be found in Extended Data Fig. 4. **d**, Kβ transient difference spectra obtained at 1-ps time delay for $[Fe(2,2'\text{-bipyridine})_3]^{2+}$ (black circles), which closely matches the model complex difference spectra (red) obtained when subtracting the singlet from the quintet spectra shown in **a**.

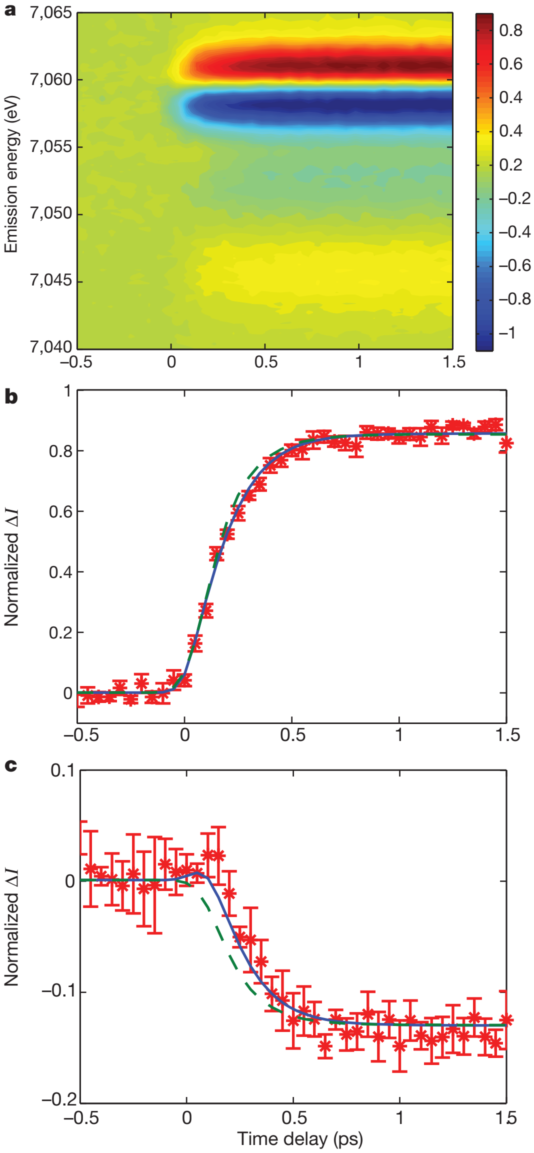


Figure 3 | Time-dependent photo-induced iron Kβ difference spectra and kinetic modelling of spin crossover dynamics. **a**, Time-dependent optically-induced two-dimensional Kβ fluorescence difference spectra for $[Fe(2,2'\text{-bipyridine})_3]^{2+}$. **b**, **c**, The difference signal measured at a Kβ fluorescence energy of 7,061 eV (**b**) and 7,054 eV (**c**) for $[Fe(2,2'\text{-bipyridine})_3]^{2+}$ (red stars), as well as the best fit achieved for kinetic models with (blue) or without (green dashed) a $^3T_{1,2}$ transient. The error bars in **b** and **c** reflect the standard error for the difference signal determined from six independent measurements.

dynamics implies that the Kβ spectra do not depend significantly on the time-evolving nuclear structure, consistent with the insensitivity of the ground-state Kβ spectra to the ligand details.

The successful analysis of the experimental data relies on two constraints presented by the model spectra shown in Fig. 2b and two constraints derived from the kinetic models. We force (1) the shape and (2) the relative amplitudes of the difference signals for the ^{1,3}MLCT, ^{3}T and $^5T_2$ electronic excited states to match the shape and relative amplitudes of the model complex difference spectra. We also require (3) all X-ray fluorescence energies to be fitted with a single time zero and (4) all MLCT excited states to undergo spin crossover, consistent with previous measurements of the spin crossover quantum yield[13]. The ultrafast rise of the difference signal shown in Fig. 3b greatly constrains the value of time zero and the final $^5T_2$ state population. For the fit to the direct spin crossover mechanism shown in Fig. 3b, the fast rise in signal at 7,061 eV requires a fast rise in $^5T_2$ population. As shown in Fig. 3c, the fast rise in the direct mechanism fit at 7,061 eV also leads to a fast drop in signal at 7,054 eV, because the $^5T_2$ state has a negative difference signal at 7,054 eV. For the fit to the sequential spin crossover mechanism also shown in Fig. 3b, the fast rise in signal at 7,061 eV can be accommodated initially by a rise in ^{3}T population. Because the ^{3}T state does not have a negative difference signal at 7,054 eV, the fast rise in ^{3}T population does not lead to a fast drop at 7,054 eV. The stepwise transition through the ^{3}T leads to a delayed onset of the drop in fluorescence amplitude at 7,054 eV relative to the rise in signal at 7,061 eV, consistent with the experimental data. For the direct model, a shift in time zero to fit the data in Fig. 3c would lead to a poor fit of the data in Fig. 3b.

Relaxation to the $^5T_2$ excited state via a ^{3}T transient provides a more satisfying explanation for the relaxation dynamics. We speculate that the sequential relaxation occurs more promptly than the direct crossover from the ^{1,3}MCLT to the $^5T_2$ excited state because the sequential transition involves single electronic transitions coupled by a spin–orbit operator, whereas the direct transition involves the simultaneous transition of two distinct electrons on two centres and cannot occur with the first-order spin–orbit operator. The sequential relaxation, like the direct transition, provides an energetically feasible pathway with minimal reaction barriers between states that can be coupled with standard spin–orbit interactions[22]. The spin–orbit matrix elements in conjunction with the calculated potential energies of a variety of electronic excited states of $[Fe(2,2'\text{-bipyridine})_3]^{2+}$ as a function of the metal–ligand bond distance provide an approximate explanation for the fast intersystem crossing and the extremely short lifetime of the ^{3}T excited state. A diagram of these potential energy surfaces can be found in Fig. 1e. In principle, the triplet ligand field excited state could be either a $^3T_1$ or a $^3T_2$ state. Computations indicate a crossing of the $^3T_2$ state in the Franck–Condon region of the ^{1,3}MLCT excited state and that the $^{1,3}\text{MLCT}\rightarrow{}^3T_2\rightarrow{}^5T_2$ pathway dominates[27]; however, relaxation trajectories involving the $^3T_1$ ligand field excited state remain plausible, and more definite conclusions will require a more complete calculation of the multidimensional potential energy surfaces, including the potentially important role of metal–ligand torsional motion[28]. The sequential model fit in Fig. 3 gives a 150 ± 50 fs time constant for ^{1,3}MLCT decay to the ^{3}T state and a 70 ± 30 fs time constant for ^{3}T decay to the $^5T_2$ state. Although the mechanistic conclusions we have drawn from our measurements differ from the earlier interpretation[26], our experimental findings do not contradict the earlier results, but rather expand on them. The extracted decay time for the ^{1,3}MLCT excited state and the effective rise time for the $^5T_2$ excited state agree with the time constants observed previously within experimental error[26]. The similarity of the ^{1,3}MLCT decay time and the $^5T_2$ rise time results from the rate of ^{3}T decay being greater than that of ^{3}T formation. This inhibits the build-up of molecules in the ^{3}T excited state and challenges the temporal differentiation of the distinct electronic states involved in spin crossover (see Extended Data Fig. 2d). Only with a technique highly sensitive to the iron spin multiplicity can the presence of the ^{3}T transient excited state in the relaxation dynamics be robustly resolved.

The complex excited-state electronic structure of molecules containing transition metals has inhibited the unambiguous interpretation of experimental measurements and the development of excited-state quantum dynamics simulations. We have demonstrated here that ultrafast X-ray fluorescence spectroscopy enables robust measurements of the charge and spin dynamics integral to excited-state relaxation in 3*d* transition-metal coordination complexes, which represents an important step towards an incisive mechanistic understanding of excited-state dynamics in 3*d* transition metal complexes.

METHODS SUMMARY

We performed femtosecond hard X-ray fluorescence measurements on a 50 mM solution of electronically excited $[Fe(2,2'\text{-bipyridine})_3]^{2+}$ in water at the XPP instrument at the LCLS. The experiment used a 0.1-mm-thick planar liquid jet oriented at 45° relative to the direction of the incident X-ray beam. The sample solution was collinearly excited with a 70-fs FWHM 520-nm laser beam. The absorption spectrum

and laser power dependence can be found in Extended Data Fig. 5. A cylindrically bent energy dispersive X-ray emission spectrometer and a 2D pixel array detector (PAD) were used to capture the iron $3p$–$1s$ (Kβ) fluorescence. The PAD response calibration involved a pixel-dependent dark current subtraction, a common mode off-set, and an experimentally determined gain correction. The final Kβ fluorescence spectrum for each time-step was obtained by integrating the signal in the non-dispersive direction. The shot-to-shot X-ray–optical relative time of arrival fluctuations were measured with a timing diagnostic and used to sort each shot by its relative time of arrival. We measured the Kβ fluorescence spectra of a series of iron model complexes with different spin states at beamline 6-2 of SSRL. We have used electronic ground-state spectra and kinetic models, with and without triplet transients, to analyse the time evolution of the Kβ fluorescence spectra.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 7 May 2013; accepted 6 March 2014.

Published online 7 May 2014.

1. Gust, D., Moore, T. A. & Moore, A. L. Mimicking photosynthetic solar energy transduction. *Acc. Chem. Res.* **34,** 40–48 (2001).
2. Sato, O., Iyoda, T., Fujishima, A. & Hashimoto, K. Photoinduced magnetization of a cobalt-iron cyanide. *Science* **272,** 704–705 (1996).
3. Ferrere, S. & Gregg, B. A. Photosensitization of TiO_2 by Fe-II(2,2′-bipyridine-4,4′-dicarboxylic acid)$_2$(CN)$_2$: band selective electron injection from ultra-short-lived excited states. *J. Am. Chem. Soc.* **120,** 843–844 (1998).
4. Heyduk, A. F. & Nocera, D. G. Hydrogen produced from hydrohalic acid solutions by a two-electron mixed-valence photocatalyst. *Science* **293,** 1639–1641 (2001).
5. Goldbeck, R. A., Kim-Shapiro, D. B. & Kliger, D. S. Fast natural and magnetic circular dichroism spectroscopy. *Annu. Rev. Phys. Chem.* **48,** 453–479 (1997).
6. Bressler, C. *et al.* Femtosecond XANES study of the light-induced spin crossover dynamics in an iron(II) complex. *Science* **323,** 489–492 (2009).
7. Huse, N. *et al.* Femtosecond soft X-ray spectroscopy of solvated transition-metal complexes: deciphering the interplay of electronic and structural dynamics. *J. Phys. Chem. Lett.* **2,** 880–884 (2011).
8. Gutlich, P. & Goodwin, H. A. in *Spin Crossover in Transition Metal Compounds I* Vol. 233, *Topics in Current Chemistry* (eds Gutlich, P. & Goodwin, H. A.) 1–47 (Springer, 2004).
9. Creutz, C., Chou, M., Netzel, T. L., Okumura, M. & Sutin, N. Lifetimes, spectra, and quenching of the excited-states of polypyridine complexes of iron(II), ruthenium(II), and osmium(II). *J. Am. Chem. Soc.* **102,** 1309–1319 (1980).
10. Hauser, A. Intersystem crossing in the Fe(PTZ)$_6$ (BF$_4$)$_2$ spin crossover system (PTZ = 1-propyltetrazole). *J. Chem. Phys.* **94,** 2741–2748 (1991).
11. McCusker, J. K. *et al.* Subpicosecond ^{1}MLCT-$^5T_2$ intersystem crossing of low-spin polypyridyl ferrous complexes. *J. Am. Chem. Soc.* **115,** 298–307 (1993).
12. Monat, J. E. & McCusker, J. K. Femtosecond excited-state dynamics of an iron(II) polypyridyl solar cell sensitizer model. *J. Am. Chem. Soc.* **122,** 4092–4097 (2000).
13. Gawelda, W. *et al.* Ultrafast nonadiabatic dynamics of $[Fe(II)(bpy)_3]^{2+}$ in solution. *J. Am. Chem. Soc.* **129,** 8199–8206 (2007).
14. Consani, C. *et al.* Vibrational coherences and relaxation in the high-spin state of aqueous $[Fe\text{-}II(bpy)_3]^{2+}$. *Angew. Chem. Int. Edn* **48,** 7184–7187 (2009).
15. Lemke, H. T. *et al.* Femtosecond X-ray absorption spectroscopy at a hard X-ray free electron laser: application to spin crossover dynamics. *J. Phys. Chem. A* **117,** 735–740 (2013).
16. Emma, P. *et al.* First lasing and operation of an angstrom-wavelength free-electron laser. *Nature Photon.* **4,** 641–647 (2010).
17. Harmand, M. *et al.* Achieving few-femtosecond time-sorting at hard X-ray free-electron lasers. *Nature Photon.* **7,** 215–218 (2013).
18. Haldrup, K. *et al.* Guest-host interactions investigated by time-resolved X-ray spectroscopies and scattering at MHz rates: solvation dynamics and photoinduced spin transition in aqueous $[Fe(bipy)_3]^{2+}$. *J. Phys. Chem. A* **116,** 9878–9887 (2012).
19. Vankó, G. *et al.* Probing the 3d spin momentum with X-ray emission spectroscopy: the case of molecular-spin transitions. *J. Phys. Chem. B* **110,** 11647–11653 (2006).
20. Krause, M. O. & Oliver, J. H. Natural widths of atomic K and L levels, K-alpha X-ray lines and several KLL Auger lines. *J. Phys. Chem. Ref. Data* **8,** 329–338 (1979).
21. de Graaf, C. & Sousa, C. Study of the light-induced spin crossover process of the $[Fe(II)(bpy)_3]^{2+}$ complex. *Chemistry* **16,** 4550–4556 (2010).
22. de Graaf, C. & Sousa, C. On the role of the metal-to-ligand charge transfer states in the light-induced spin crossover in Fe-II(bpy)$_3$. *Int. J. Quantum Chem.* **111,** 3385–3393 (2011).
23. Glatzel, P. & Bergmann, U. High resolution 1s core hole X-ray spectroscopy in 3d transition metal complexes — electronic and structural information. *Coord. Chem. Rev.* **249,** 65–95 (2005).
24. de Groot, F. High resolution X-ray emission and X-ray absorption spectroscopy. *Chem. Rev.* **101,** 1779–1808 (2001).
25. Lee, N., Petrenko, T., Bergmann, U., Neese, F. & DeBeer, S. Probing valence orbital composition with iron K β X-ray emission spectroscopy. *J. Am. Chem. Soc.* **132,** 9715–9727 (2010).
26. Cannizzo, A. *et al.* Light-induced spin crossover in Fe(II)-based complexes: the full photocycle unraveled by ultrafast optical and X-ray spectroscopies. *Coord. Chem. Rev.* **254,** 2677–2686 (2010).
27. Sousa, C. *et al.* Ultrafast deactivation mechanism of the excited singlet in the light-induced spin crossover of $[Fe(2,2'\text{-bipyridine})_3]^{2+}$. *Chemistry* **19,** 17541–17551 (2013).
28. Alvarez, S. Relationships between temperature, magnetic moment, and continuous symmetry measures in spin crossover complexes. *J. Am. Chem. Soc.* **125,** 6795–6802 (2003).

Acknowledgements We thank P. Frank, B. Lin and S. DeBeer for discussion, S. DeBeer for some model iron complex X-ray fluorescence spectra, and D. Stanbury for providing some iron complexes. Experiments were carried out at LCLS and SSRL, which are National User Facilities operated for DOE and OBES respectively by Stanford University. W.Z., R.W.H., H.W.L., D.A.M., Z.S. and K.J.G. acknowledge support from the AMOS programme within the Chemical Sciences, Geosciences and Biosciences Division of the Office of Basic Energy Sciences, Office of Science, US Department of Energy. E.I.S. acknowledges support from the NSF (CHE-0948211). R.G.H. acknowledges a Gerhard Casper Stanford Graduate Fellowship and the Achievements Rewards for College Scientists (ARCS) Foundation. T.K. acknowledges the German Research Foundation (DFG), grant KR3611/2-1. K.S.K., M.M.N. and T.B.v.D. acknowledge support from the Danish National Research Foundation and from DANSCATT. K.K. thanks the Volkswagen Foundation for support under the Peter Paul Ewald fellowship program (I/85832). G.V. acknowledges support from the European Research Council (ERC-StG-259709) and the Lendület Programme of the Hungarian Academy of Sciences. C.B., W.G. and A.G. thank the DFG (SFB925), as well as the European XFEL, for financial support.

Author Contributions W.Z., R.A.-M., U.B., R.W.H., D.A.M., T.-C.W. and K.J.G. designed the experiment. W.Z., R.A.-M., U.B., M.C., R.W.H., K.S.K., K.K., H.T.L., H.W.L., C.P., J.S.R., Z.S., D.S., T.B.v.D., T.-C.W., D.Z. and K.J.G. did the experiment. W.Z., T.K., K.S.K., T.B.v.D., G.V. and T.-C.W. analysed the data. W.Z., R.A.-M., U.B., C.B., W.G., A.G., R.G.H., R.W.H., T.K., K.S.K., K.K., D.A.M., M.M.N., E.I.S., D.S. and K.J.G. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.J.G. (kgaffney@slac.stanford.edu).

METHODS

Experimental procedures. We performed femtosecond hard X-ray fluorescence measurements on a 50 mM solution of $[Fe(2,2'\text{-bipyridine})_3]^{2+}$ in water at the X-ray pump-probe (XPP) instrument at the Linac Coherent Light Source (LCLS). The experiment used a 0.1 mm thick planar liquid jet oriented at an angle of 45° with respect to the direction of the incident X-ray beam. We measured the ultraviolet–visible absorption spectrum of the solution before and after the measurement to ensure no appreciable sample damage had occurred. The sample solution was collinearly excited with a 70 fs FWHM 520 nm laser beam (120 mJ cm^{-2}) generated by optical parametric amplification of the 800 nm output of a Ti:sapphire regenerative amplifier laser system (Coherent, Legend). With 520 nm light, we excited $[Fe(2,2'\text{-bipyridine})_3]^{2+}$ at the peak of the MLCT band (Extended Data Fig. 5a). We set the pump laser fluence to maximize excitation yield, while avoiding other deleterious photophysical phenomena. Previous time-resolved hard X-ray spectroscopy measurements of iron spin crossover compounds have used higher, often significantly higher, optical laser fluence[29–31]. We used an excitation laser fluence where the transient optical signal changes linearly with pump fluence, as shown in Extended Data Fig. 5b. The 8 keV X-ray laser pulses, with an average bandwidth of 0.3%, were focused using Be compound refractive lenses to a 50 μm diameter spot size at the sample position. Shot-to-shot fluctuations in the X-ray incidence energy and band width do not influence the X-ray fluorescence spectrum when the X-ray energy is well above the core ionization threshold. For iron, with a 1*s* ionization threshold of 7.112 keV, the 8 keV X-ray energy used in the experiment achieves this goal.

The incoming X-ray pulse energy was measured using non-invasive diagnostics before the sample[32]. A high-resolution energy dispersive X-ray emission spectrometer[33], based on the von Hamos geometry, was used to capture the iron 3*p*–1*s* (Kβ) fluorescence. The spectrometer was equipped with 4 cylindrically bent (0.5 m radius) Ge(620) crystal analysers and set to cover a Bragg angle range from 78.0° to 80.4°. The CSPAD 2D pixel array detector (388 × 370 pixels)[34] intersected the X-rays diffracted from the crystal analysers in an energy range from 7,033 to 7,084 eV.

The detector response calibration involved a pixel dependent dark current (pedestal) subtraction, a common mode offset, and an experimentally determined gain map. The gain map was built from histograms of each pixel response extracted from multiple images (after dark current and common mode offset corrections) collected over many minutes. Gaussians were fitted to the zero and one photon peaks of the histograms, enabling fine-tuned dark and gain corrections to the histograms directly from the data. The zero photon peaks were centred at zero analogue-to-digital units and the separation between the zero and one photon peaks were scaled to unity for all pixels. The counts for each pixel in a given time-step were obtained by averaging the analogue-to-digital values above a threshold of 2.5σ of the zero-photon peak and scaling to the incident X-ray intensity. The final 1D spectrum for each time-step was obtained by integrating the signal in the non-dispersive direction[33].

The shot-to-shot X-ray–optical relative time of arrival fluctuations were measured for every X-ray–optical pulse pair with a timing diagnostic tool based on optical detection of X-ray generated carriers in a Si_3N_4 thin film. A description of the time diagnostic tool and the demonstrated performance of the tool can be found elsewhere[17,35]. This experimental measure of the relative timing can be used to sort each experimental shot by the relative time of arrival. Although the timing tool provides an accurate measure of the shot-to-shot variation in the relative time of arrival between the X-ray and optical laser pulses, it does not provide an accurate measure of the instrument response function. The timing tool uses changes in the Si_3N_4 dielectric function to modify the transmission of a chirped white light pulse through the Si_3N_4 thin film. These changes in the dielectric function result from the increase in free carriers generated by X-ray ionization, Auger relaxation and impact ionization. The temporal response is the convolution of these complex dynamics with the cross-correlation of the X-ray and optical laser pulses. Without a detailed model of the carrier generation, the cross-correlation cannot be extracted from the timing tool. At present, no experimental means of cross-correlating the hard X-ray and optical pulses has been demonstrated.

The final time resolution of the experiment results from the convolution of the optical and X-ray pulse durations, the group velocity walk-off of the X-ray and optical pulses in the sample and the error in the relative time of arrival measurement. These factors would predict a cross-correlation of roughly 150 fs FWHM. In the data analysis, the instrument response function FWHM and time zero (coincident arrival of the X-ray and optical pulses) are fit parameters.

Kβ fluorescence spectra for model complexes. The 3*p*–1*s* X-ray (Kβ) fluorescence spectra of model complexes play an important role in our analysis of the time-dependent data. The Kβ fluorescence spectra of 3*d* transition-metal ions reflect the 3*p*−3*d* exchange interaction, which makes the line shapes sensitive to the spin state of the transition metal atom[19,23,24,36,37]. Kβ fluorescence provides a powerful technique for spin state studies, particularly when there are advantages of working with penetrating hard X-rays. When a sample contains multiple spin states, the spin state distribution can be readily and precisely calculated from the line shape variations[19].

We measured the Kβ fluorescence spectra of a series of iron complexes with different spin states at beamline 6-2 of the Stanford Synchrotron Radiation Lightsource (SSRL). All the samples were cooled to 10 K to reduce the influence of X-ray damage. The static spectra, collected with a multi-crystal high-resolution X-ray emission spectrometer, are shown in Fig. 2a.

We use the model complex difference spectra generated from molecules that have different spin multiplicities in their electronic ground state to model the time-dependent populations of electron excited states with different spin multiplicities. We verify the validity of using the model complex difference spectra generated from the quintet $[Fe(phenanthroline)_2(NCS)_2]$ and the singlet $[Fe(2,2'\text{-bipyridine})_3]^{2+}$ model compounds for the quintet excited state by comparing it with the transient difference spectra of $[Fe(2,2'\text{-bipyridine})_3]^{2+}$ after a 1-ps time delay (see Fig. 2d). The validity of model complex difference spectra for the ^{1,3}MLCT and ^{3}T excited states proves more challenging to demonstrate because we do not isolate these excited states at any time delay in our pump probe measurements (the fit to the 50-fs time delay spectra shown in Fig. 2c indicates a population ratio of 5:1.3:1 for the ^{1,3}MLCT:^{3}T:$^5T_2$ excited states).

Despite this limitation, the model for the ^{1,3}MLCT excited generated from doublet $[Fe(2,2'\text{-bipyridine})_3]^{3+}$ and singlet $[Fe(2,2'\text{-bipyridine})_3]^{2+}$ compounds should be robust since the only distinction is the presence of the electron on the 2,2′-bipyridine ligand which should have minimal impact on the Kβ fluorescence spectrum. For the ^{3}T transient, no long-lived triplet excited state can be used to extract an excited state Kβ fluorescence difference spectrum as an internal reference. Instead, we use the ground state model complex difference spectrum obtained from triplet Fe(II) phthalocyanine (FePc) and singlet $[Fe(2,2'\text{-bipyridine})_3]^{2+}$ Kβ spectra as our reference difference spectra. We used the four-coordinate FePc, rather than an octahedral model complex, because octahedral Fe(II) complexes cannot have a triplet ground state. While de Beer *et al.* have shown that tetrahedral, octahedral, and square planar molecules in the same quintet or sextet spin state have very similar Kβ spectra[25], this cannot be demonstrated experimentally for intermediate spin states. Instead, we use theoretical calculations to demonstrate this point. We theoretically calculated the Kβ fluorescence spectra of a four-coordinate square planar and a six-coordinate octahedral ferrous complex using atomic multiplet theory[38]. This theory is the standard method for calculating and interpreting hard X-ray fluorescence spectra[38]. For all calculations, the Slater–Condon parameters were reduced to 80% of their atomic value and the 3*d* orbital and spin angular momentum (LS) coupling was switched off for simplicity. The Kβ spectra were calculated as a 3*p*→1*s* fluorescence following 1*s* ionization. For FePc, we use the previously published crystal field parameters ($10Dq = 2.7$ eV, $Ds = 0.86$ and $Dt = 0.247$) in our calculations[39]. For the six-coordinate octahedral complex calculation, we used a $10Dq = 1.5$ eV, consistent with the experimental $10Dq \approx 1.5$ eV measured for $[Fe(2,2'\text{-bipyridine})_3]^{2+}$ (ref. 9). This value also ensures a low spin ($S = 0$) ground state, a high spin ($S = 2$) first excited state and an intermediate spin ($S = 1$) second excited state.

Extended Data Fig. 1a shows the calculated Kβ fluorescence spectra for both the four- and six-coordinate complexes. The square planar and octahedral symmetries have similar triplet state Kβ fluorescence spectra, consistent with prior experimental and theoretical findings for high spin complexes[19,25]. The accuracy of the calculations can also be assessed by comparing calculated and experimental difference spectra. In Extended Data Fig. 1b we show a comparison between the calculated difference spectrum generated when subtracting an octahedral crystal field singlet state from the square planar triplet ground state and the experimental difference spectrum generated by subtracting singlet $[Fe(2,2'\text{-bipyridine})_3]^{2+}$ spectrum from the triplet FePc spectrum. The calculated difference spectrum reproduces the qualitative features of the experimental difference spectrum. The insensitivity of the calculated spectra to the coordination geometry and the ability of the calculations to reproduce the main features of the experimental difference spectrum validate the use of the FePc fluorescence spectrum as a model for the triplet excited state of $[Fe(2,2'\text{-bipyridine})_3]^{2+}$.

Using model complex difference spectra has proven more fruitful for the kinetic modelling than singular value decomposition (SVD). The model complex difference spectra demonstrate that differentiation of the ^{1,3}MLCT and the ^{3}T excited states depends upon both the shape of the difference spectra and the relative amplitudes of the difference spectra. To first order, the integrated area of the Kβ fluorescence spectra do not change with spin state. The integral of the absolute value of the difference spectrum, however, depends linearly on the magnitude of the spin change[39]. This robust and reproducible aspect of Kβ fluorescence spectroscopy makes the relative amplitudes of the difference spectra an important distinction. SVD, however, struggles to differentiate species when a difference in relative amplitude is a key distinguishing feature of the difference spectra. For this reason, we have used model complex difference spectra, rather than SVD to model the time resolved data.

Kinetic modelling of the $[Fe(2,2'\text{-bipyridine})_3]^{2+}$ experimental population dynamics. We have used two distinct kinetic models to analyse the time-dependent electron dynamics in $[Fe(2,2'\text{-bipyridine})_3]^{2+}$. For the direct transition between ^{1,3}MLCT

and $^5T_2$, without a ^3T transient state, the relaxation mechanism can be expressed as follows:

$$^{1,3}\mathrm{MLCT} \xrightarrow{k_1} {}^5\mathrm{T}_2 \xrightarrow{k_3} {}^1\mathrm{A}_1$$

where ^{1,3}MLCT corresponds to the electronic excited state populated by optical excitation, $^5T_2$ corresponds to the quintet ligand field state, and $^1A_1$ represents the electronic ground state. The differential rate equations for each species are given by the following mass balance simultaneous equations,

$$\frac{d[^{1,3}\mathrm{MLCT}]}{dt} = -k_1[^{1,3}\mathrm{MLCT}]$$

$$\frac{d[^5\mathrm{T}_2]}{dt} = k_1[^{1,3}\mathrm{MLCT}] - k_3[^5\mathrm{T}_2]$$

$$\frac{d[^1\mathrm{A}_1]}{dt} = k_3[^5\mathrm{T}_2]$$

The integrated rate equations provide the following time-dependent populations for the three species,

$$[^{1,3}\mathrm{MLCT}] = [^{1,3}\mathrm{MLCT}]_0 e^{-k_1 t}$$

$$[^5\mathrm{T}_2] = [^{1,3}\mathrm{MLCT}]_0 \frac{k_1}{k_3 - k_1}(e^{-k_1 t} - e^{-k_3 t})$$

$$[^1\mathrm{A}_1] = [^{1,3}\mathrm{MLCT}]_0 - [^{1,3}\mathrm{MLCT}] - [^5\mathrm{T}_2]$$

From prior ultrafast measurements, we know that the lifetime of the $^5T_2$ excited state is roughly 660 ps (refs 6, 13, 15). The long lifetime of the $^5T_2$ excited state allows us to set $k_3 \approx 0$ when we model the kinetics in the first couple of picoseconds. The integrated rate equations can be reduced to:

$$[^{1,3}\mathrm{MLCT}] = [^{1,3}\mathrm{MLCT}]_0 e^{-k_1 t}$$

$$[^5\mathrm{T}_2] = [^{1,3}\mathrm{MLCT}]_0 (1 - e^{-k_1 t})$$

For the sequential kinetic model with a ^3T transient state, the relaxation mechanism can be expressed as follows:

$$^{1,3}\mathrm{MLCT} \xrightarrow{k_1} {}^3\mathrm{T} \xrightarrow{k_2} {}^5\mathrm{T}_2 \xrightarrow{k_3} {}^1\mathrm{A}_1$$

where ^{1,3}MLCT corresponds to the electronic excited state populated by optical excitation, ^3T corresponds to the triplet ligand field excited state, and $^5T_2$ corresponds to the quintet ligand field excited state, and $^1A_1$ represents the electronic ground state. The differential rate equations for each species are given by the following mass balance simultaneous equations:

$$\frac{d[^{1,3}\mathrm{MLCT}]}{dt} = -k_1[^{1,3}\mathrm{MLCT}]$$

$$\frac{d[^3\mathrm{T}]}{dt} = k_1[^{1,3}\mathrm{MLCT}] - k_2[^3\mathrm{T}]$$

$$\frac{d[^5\mathrm{T}_2]}{dt} = k_2[^3\mathrm{T}] - k_3[^5\mathrm{T}_2]$$

$$\frac{d[^1\mathrm{A}_1]}{dt} = k_3[^5\mathrm{T}_2]$$

The integrated rate equations provide the following time-dependent populations for the four species:

$$[^{1,3}\mathrm{MLCT}] = [^{1,3}\mathrm{MLCT}]_0 e^{-k_1 t}$$

$$[^3\mathrm{T}] = [^{1,3}\mathrm{MLCT}]_0 \frac{k_1}{k_2 - k_1}(e^{-k_1 t} - e^{-k_2 t})$$

$$[^5\mathrm{T}_2] = [^{1,3}\mathrm{MLCT}]_0 \frac{k_1 k_2[(k_3 - k_2)e^{-k_1 t} - (k_3 - k_1)e^{-k_2 t} + (k_2 - k_1)e^{-k_3 t}]}{(k_2 - k_1)(k_3 - k_2)(k_3 - k_1)}$$

$$= [^{1,3}\mathrm{MLCT}]_0 \frac{k_1 k_2[k_3(e^{-k_1 t} - e^{-k_2 t}) + k_2(e^{-k_3 t} - e^{-k_1 t}) + k_1(e^{-k_2 t} - e^{-k_3 t})]}{(k_2 - k_1)(k_3 - k_2)(k_3 - k_1)}$$

$$[^1\mathrm{A}_1] = [^{1,3}\mathrm{MLCT}]_0 - [^{1,3}\mathrm{MLCT}] - [^3\mathrm{T}] - [^5\mathrm{T}_2]$$

The long lifetime of the $^5T_2$ excited state allows us to set $k_3 \approx 0$ when we model the kinetics in the first couple of picoseconds. The integrated rate equations can be reduced to:

$$[^{1,3}\mathrm{MLCT}] = [^{1,3}\mathrm{MLCT}]_0 e^{-k_1 t}$$

$$[^3\mathrm{T}] = [^{1,3}\mathrm{MLCT}]_0 \frac{k_1}{k_2 - k_1}(e^{-k_1 t} - e^{-k_2 t})$$

$$[^5\mathrm{T}_2] = [^{1,3}\mathrm{MLCT}]_0 (1 - \frac{k_2 e^{-k_1 t} - k_1 e^{-k_2 t}}{k_2 - k_1})$$

To fit the experimental data to a kinetic model, we must convolve the kinetic model with the instrument response function which we describe with a Gaussian function. Taking the example of $[^{1,3}\mathrm{MLCT}] = [^{1,3}\mathrm{MLCT}]_0 e^{-k_1 t}$, which is an exponential decay starting at time zero (t_0), it will be expressed as

$$[^{1,3}\mathrm{MLCT}] = [^{1,3}\mathrm{MLCT}]_0 \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-y^2/2\sigma^2} \mathrm{H}(t - t_0 - y) e^{-k_1(t - t_0 - y)} dy$$

where H is the Heaviside step function and σ is the temporal width of the instrument response function.

Statistical determination of the correct kinetic model. Given two distinct kinetic models, we must determine which model best represents the experimental data. Choosing the model with smaller residual sum of squares (RSS) is not sufficient because the two models do not have the same number of fit parameters. We have used the statistical F-test to determine whether the model with or without a ^3T transient provides the best fit of the experimental data[40].

The F-test provides a statistically robust method for comparing the quality of two models with a different number of fit parameters when the simpler model 1 can be 'nested' within the more complex model 2. Model 1 has p_1 parameters, and model 2 has p_2 parameters, where $p_2 > p_1$. For any choice of parameters in model 1, the model 2 should always be able to fit the data at least as well as the model 1. Thus, model 2 typically will have a lower RSS than model 1. The F-test allows us to determine the statistical significance of this variance in RSS. The F statistic can be calculated by

$$F = \frac{\dfrac{\mathrm{RSS}_1 - \mathrm{RSS}_2}{p_2 - p_1}}{\dfrac{\mathrm{RSS}_2}{n - p_2}} = \frac{(\mathrm{RSS}_1 - \mathrm{RSS}_2)(n - p_2)}{\mathrm{RSS}_2(p_2 - p_1)}$$

where n is the number of data points (time delays) fitted by the two models. For the null hypothesis that model 2 does not provide a fit statistically superior to that provided by model 1, the F will have an F distribution defined by the degrees of freedom, $(p_2 - p_1)$ and $(n - p_2)$. To reject the null hypothesis, F must exceed a critical value that depends upon the degrees of freedom and the level of confidence[40].

$[Fe(2,2'\text{-bipyridine})_3]^{2+}$ experimental data modelling. Using the reference difference spectra with the kinetic model, we fit the time-dependent difference Kβ fluorescence spectra for optically excited $[Fe(2,2'\text{-bipyridine})_3]^{2+}$ in water. The parameters extracted from the fit of the two kinetic models can be found in Extended Data Table 1. We compute the time constants and uncertainties reported in Extended Data Table 1 by fitting multiple runs of the same experiment and then calculating the mean and the standard deviation. The experimental two-dimensional transient difference spectra, fit spectra, residuals, and excited electronic state populations extracted from the best fit for each model can be found in Extended Data Figs 2 and 3. Given the very short lifetime of the ^3T excited state, the deviations between the fits of the two models predominantly occur within the first 500 fs. The two-dimensional plots of the residuals in Extended Data Figs 2c and 3c highlight the regions where the $^{1,3}\mathrm{MLCT}\rightarrow{}^3\mathrm{T}\rightarrow{}^5\mathrm{T}_2$ model provides a fit superior to that of the $^{1,3}\mathrm{MLCT}\rightarrow{}^5\mathrm{T}_2$ model. Unsurprisingly, this corresponds to time delays with larger ^3T populations and spectral regions with the largest difference between the ^3T and $^5T_2$ spectra (7,053–7,056 eV).

The residual sum of squares quantifies the variable quality in the fits. The residual sum of squares for each model is: $\mathrm{RSS}_1 = 3.77$ and $\mathrm{RSS}_2 = 3.21$. In this situation, we have $p_1 = 5$, $p_2 = 6$ and $n = 45$. To be 95% confident that the complex model is better than the nested model, the calculated F value must be larger than the F distribution value that captures 95% of the distribution for $F(p_2 - p_1, n - p_2)$ which is 4.09. The calculated F value is 6.71 which exceeded 4.09. So with 95% confidence we conclude that the model containing the ^3T transient provides a better description of the experimental data.

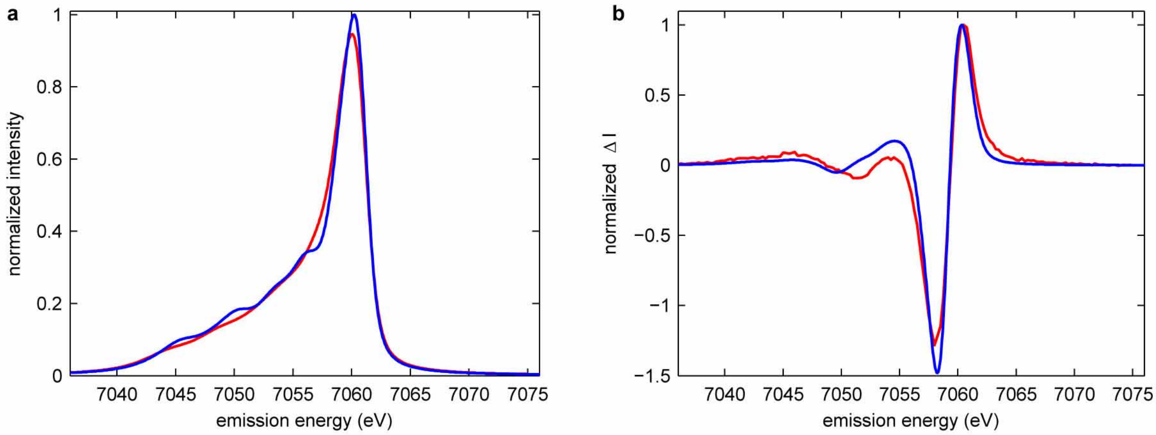
Influence of instrument response function parameters on the data analysis. We utilize the instrument response function (IRF) as a variable since the technology does not yet exist to measure the instrument response time accurately. This leads to an increase in the number of parameters in the data analysis. This increase in fit parameters makes statistically differentiating the robustness of alternative kinetic models more difficult, rather than easier.

To ensure that the statistical superiority of the kinetic model possessing the ^3T transient does not result from our uncertainty about the instrument response function parameters, we have investigated how variation of time zero and FWHM values differentially influence the RSS for the direct $^{1,3}\mathrm{MLCT}\rightarrow{}^5\mathrm{T}_2$ model and the $^{1,3}\mathrm{MLCT}\rightarrow{}^3\mathrm{T}\rightarrow{}^5\mathrm{T}_2$ model. For the range of time zero and FWHM values reported in Extended Data Table 1 that adequately fit the experimental data with either model, the model containing the ^3T transient always provides a significantly superior fit to the experimental data. We have used the instrument response function values that minimize the RSS for the $^{1,3}\mathrm{MLCT}\rightarrow{}^5\mathrm{T}_2$ model to fit the data with the $^{1,3}\mathrm{MLCT}\rightarrow{}^3\mathrm{T}\rightarrow{}^5\mathrm{T}_2$ model. Using this sub-optimal instrument response function only increases the RSS_2 from 3.21 to 3.27, both significantly less than the direct model $\mathrm{RSS}_1 = 3.77$. Using the definition for F given above and $p_1 = 5$, $p_2 = 6$ and

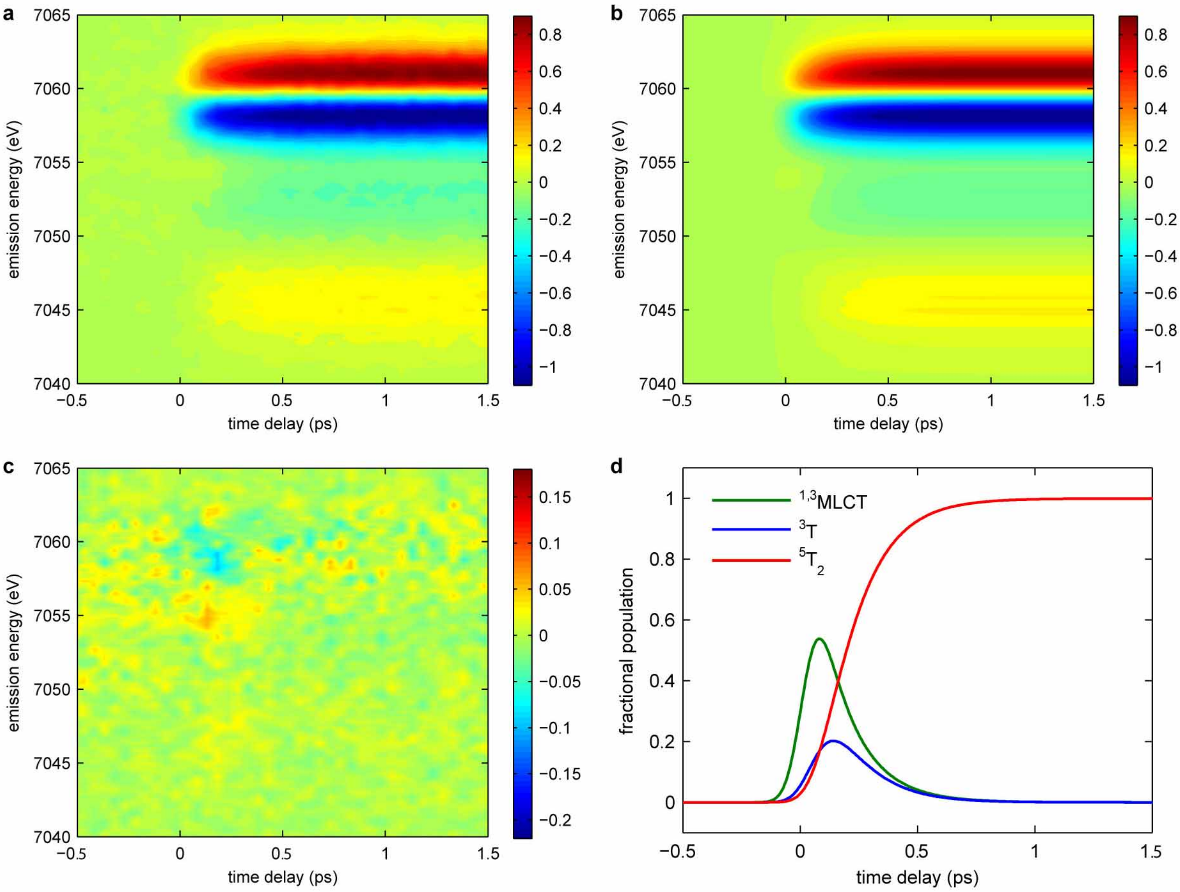
$n = 45$, we calculate $F = 5.98$, in excess of the 4.09 value needed to conclude with 95% confidence that the complex model provides a better representation of the experimental data than the nested model.

Experimental time resolution can also influence the ability to identify a distinct excited state. For the case of the triplet transient, the temporal resolution of 150 fs has little impact on the characterization of the triplet excited state dynamics. To demonstrate that the roughly 150 fs FWHM IRF does not inhibit our ability to characterize the triplet population dynamics, we have simulated the ^{1,3}MLCT$\rightarrow$^{3}T$\rightarrow$$^{5}T_2$ population kinetics using the time constants extracted from the best fit to the experimental data listed in Extended Data Table 1 with an IRF possessing a 150 fs FWHM and a 5 fs FWHM. The initial time dependence of the ^{1,3}MLCT state signal depends significantly on the time resolution (though the decays for time delays longer than 200 fs look similar), but the shape and amplitude of the triplet population is similar. The convolution of the IRF and the lifetime of the ^{1,3}MLCT excited state determine the time dependence of the ^{3}T transient state formation observed experimentally. The low transient population of the triplet state results primarily from the fact that the decay rate of the ^{3}T state exceeds that of the ^{1,3}MLCT state by a factor of two.

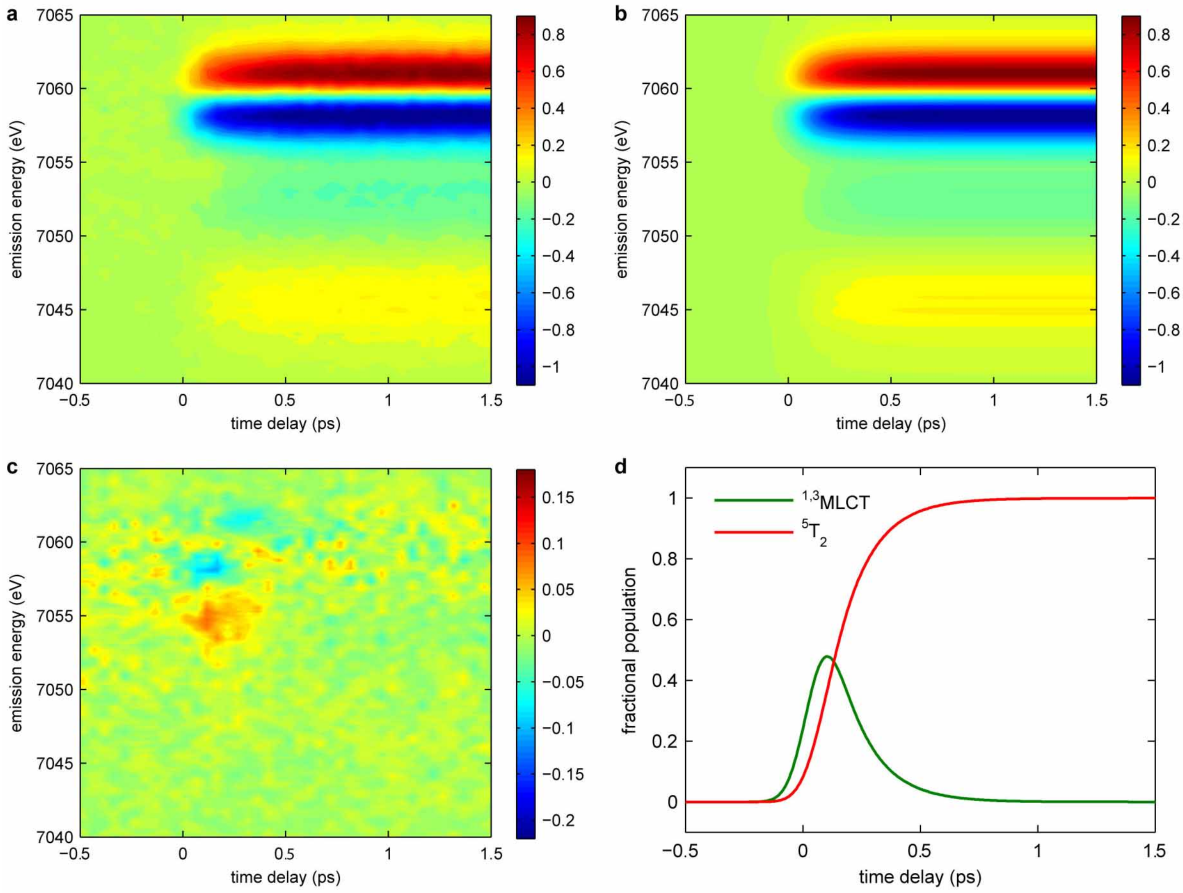
29. Khalil, M. *et al.* Picosecond X-ray absorption spectroscopy of a photoinduced iron(II) spin crossover reaction in solution. *J. Phys. Chem. A* **110,** 38–44 (2006).
30. Nozawa, S. *et al.* Direct probing of spin state dynamics coupled with electronic and structural modifications by picosecond time-resolved XAFS. *J. Am. Chem. Soc.* **132,** 61–63 (2010).
31. Gawelda, W. *Time-Resolved X-Ray Absorption Spectroscopy of Transition Metal Complexes.* Ph.D. thesis, École Polytechnique Fédérale de Lausanne (2006).
32. Feng, Y. P. *et al.* A single-shot intensity-position monitor for hard X-ray FEL sources. *Proc. SPIE* **8140,** 81400Q (2011).
33. Alonso-Mori, R. *et al.* A multi-crystal wavelength dispersive X-ray spectrometer. *Rev. Sci. Instrum.* **83,** 9 (2012).
34. Koerner, L. J., Philipp, H. T., Hromalik, M. S., Tate, M. W. & Gruner, S. M. X-ray tests of a pixel array detector for coherent X-ray imaging at the Linac Coherent Light Source. *J. Instrum.* **4,** P03001 (2009).
35. Bionta, M. R. *et al.* Spectral encoding of X-ray/optical relative delay. *Opt. Express* **19,** 21855–21865 (2011).
36. Vankó, G. *et al.* Picosecond time-resolved X-ray emission spectroscopy: ultrafast spin-state determination in an iron complex. *Angew. Chem. Int. Edn* **49,** 5910–5912 (2010).
37. Vankó, G. *et al.* Spin-state studies with XES and RIXS: From static to ultrafast. *J. Elec. Spec. Relat. Phenom.* **188,** 166–171 (2013).
38. de Groot, F. M. F. & Kotani, A. *Core Level Spectroscopy of Solids* (CRC Press, Boca Raton, 2008).
39. Stepanow, S. *et al.* Mixed-valence behavior and strong correlation effects of metal phthalocyanines adsorbed on metals. *Phys. Rev. B* **83,** 220401 (2011).
40. Kutner, M. H., Nachtsheim, C. J. & Neter, J. *Applied Linear Regression Models* (McGraw-Hill/Irwin, 2004).



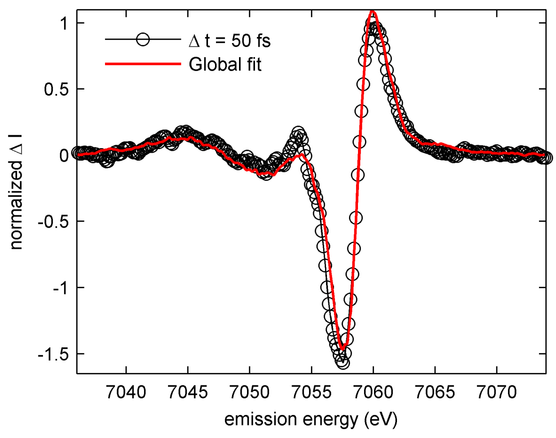
Extended Data Figure 1 | Experimental and calculated Kβ fluorescence spectra for triplet spin states. **a**, The calculated Kβ fluorescence spectra of iron complexes: triplet Fe(II) in square planar crystal field (red) (calculation parameters based on Fe(II)phthalocyanine), and triplet excited state in an octahedral crystal field (blue) (calculation parameters based on $[Fe(2,2'\text{-bipyridine})_3]^{2+}$). **b**, The experimental Kβ fluorescence difference spectrum (red) obtained by subtracting the singlet $[Fe(2,2'\text{-bipyridine})_3]^{2+}$ spectrum from the triplet Fe(II)phthalocyanine spectrum, and the calculated Kβ fluorescence difference spectrum (blue) generated by subtracting the spectrum of the singlet state in an octahedral crystal field from the triplet state in a square planar crystal field.



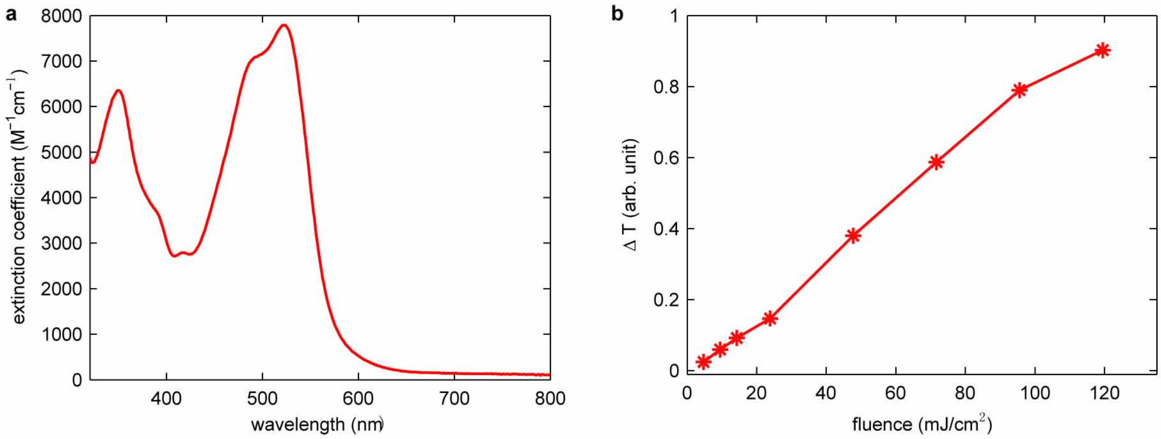
Extended Data Figure 2 | Time-dependent Kβ fluorescence spectra and fit using the sequential kinetic model with a triplet transient. **a**, Experimental transient fluorescent amplitude difference spectra plotted with arbitrary units, and **b**, fit using the sequential kinetic model with a triplet transient. **c**, Residuals for the best fit, with the colour-scale maximum and minimum set to one-fifth of the value used in **a** and **b**. **d**, The excited state populations extracted from the best fit.



Extended Data Figure 3 | Time-dependent Kβ fluorescence spectra and fit using the direct kinetic model without a triplet transient. **a**, Experimental transient fluorescent amplitude difference spectra plotted with arbitrary units, and **b**, fit using the direct kinetic model without a triplet transient. **c**, Residuals for the best fit with the colour scale maximum and minimum set to one-fifth of the value used in **a** and **b**. **d**, The excited state populations extracted from the best fit.



Extended Data Figure 4 | The 50 fs time delay normalized Kβ fluorescent amplitude difference spectrum (Δ*I*) and kinetic model fit plotted as a function of X-ray emission energy. The measured data (black circles and line), along with the best global fit from the sequential kinetic model with a transient triplet state (red line).



Extended Data Figure 5 | Absorption spectrum and pump power dependence measurements. **a**, The ultraviolet–visible absorption spectrum of $[Fe(2,2'\text{-bipyridine})_3]^{2+}$ in water. **b**, Power (fluence) dependence of the change in probe transmission measured at 520 nm, following excitation of an aqueous solution of $[Fe(2,2'\text{-bipyridine})_3]Cl_2$ with a 520 nm pump pulse. The figure shows the change in transmission (ΔT) measured at a 10 ps time delay, a time long compared to the spin crossover and vibrational cooling timescales, but short compared to the lifetime of the high-spin excited state.

Extended Data Table 1 | Fitted model parameters

Kinetic model	lifetime	lifetime	Time zero	Instrument response	
	$1/k_1$(fs)	$1/k_2$(fs)	t_0(fs)	σ(fs)	FWHM(fs)
with triplet transient	150±50	70±30	0±7	56±8	130±20
without triplet transient	140±12		15±6	70±7	170±15

Values shown are extracted from fits to sequential and direct spin crossover models for photo-excited $[Fe(2,2'\text{-bipyridine})_3]^{2+}$ in water. We compute the time constants and uncertainties by fitting six runs of the same experiment and then calculating the mean and standard deviation.

doi:10.1038/nature13278

The poleward migration of the location of tropical cyclone maximum intensity

James P. Kossin[1], Kerry A. Emanuel[2] & Gabriel A. Vecchi[3]

Temporally inconsistent and potentially unreliable global historical data hinder the detection of trends in tropical cyclone activity[1–3]. This limits our confidence in evaluating proposed linkages between observed trends in tropical cyclones and in the environment[4,5]. Here we mitigate this difficulty by focusing on a metric that is comparatively insensitive to past data uncertainty, and identify a pronounced poleward migration in the average latitude at which tropical cyclones have achieved their lifetime-maximum intensity over the past 30 years. The poleward trends are evident in the global historical data in both the Northern and the Southern hemispheres, with rates of 53 and 62 kilometres per decade, respectively, and are statistically significant. When considered together, the trends in each hemisphere depict a global-average migration of tropical cyclone activity away from the tropics at a rate of about one degree of latitude per decade, which lies within the range of estimates of the observed expansion of the tropics over the same period[6]. The global migration remains evident and statistically significant under a formal data homogenization procedure[3], and is unlikely to be a data artefact. The migration away from the tropics is apparently linked to marked changes in the mean meridional structure of environmental vertical wind shear and potential intensity, and can plausibly be linked to tropical expansion, which is thought to have anthropogenic contributions[6].

Inconsistencies in the historical global 'best-track' data can introduce substantial uncertainty into global-mean measures of tropical cyclone activity. Since the introduction of geostationary weather satellites in the mid to late 1970s, measures of tropical cyclone frequency are generally considered to be accurate, and there is no observed trend in global frequency since that time[7,8]. Comparatively, measures of tropical cyclone intensity are considered to be highly uncertain in the global data[3,9]. Consequently, storm duration is also uncertain because identifying the moment when a cyclone forms (cyclogenesis) requires accuracy in intensity estimates, as the definition of cyclogenesis is entirely dependent on a nascent storm's intensity reaching a formally specified threshold. Similar uncertainty exists in identifying a cyclone's demise (cyclolysis). These uncertainties can project onto metrics such as power dissipation[10] and accumulated cyclone energy[11], which are amalgamations of frequency, duration and intensity.

But measurements of a storm's position taken around the time that it reaches its lifetime-maximum intensity (LMI) are much less uncertain. By this time in a storm's evolution, it is more likely to have been detected and to be under close observation. Measurements of storm position at the time of LMI are also less sensitive to inaccuracy in measurements of intensity, as well as to known interregional differences in wind-averaging techniques[9], because determining the absolute LMI is not critical—it is necessary only to know that the intensity has peaked. This also makes measurements of storm position at the time of LMI comparatively insensitive to temporal heterogeneity in the historical best-track intensity record[3]. It is this heterogeneity that has presented substantial challenges to trend detection in tropical cyclone metrics that require absolute measures of intensity[5].

Here we consider the 31-yr period 1982–2012. In this period, the global best-track data are considered most complete and at their highest quality in each basin[9], and storm position is well monitored globally by geostationary satellites. This period also encompasses a recent satellite-based global tropical cyclone intensity reanalysis[3], and is the interval over which the atmospheric reanalysis products[12–14] that provide information on the environmental changes that affect tropical cyclones are most reliable.

When the annual-mean latitude of LMI is calculated from the best-track data in the Northern and Southern hemispheres over this period (Fig. 1a, b, red lines), there are clear and statistically significant poleward trends in both hemispheres of 53 and 62 km per decade, respectively (Table 1). The positive contribution to these hemispheric trends from each ocean basin except that of the North Indian Ocean (Table 1 and Extended Data Fig. 1) suggests that the migration away from the tropics is a global phenomenon, although there are large regional differences in the trend amplitudes and their statistical power. These differences are probably due, in part, to regional differences in interannual to multidecadal variability[15]. The largest contribution to the Northern Hemisphere trend is from the western North Pacific Ocean, which is also the most active basin in terms of annual tropical cyclone frequency. By contrast, the North Indian Ocean has the lowest mean annual frequency, and the small equatorward trend there has a much lesser effect on the hemispheric trend. The North Atlantic Ocean and eastern North Pacific exhibit small poleward trends and also contribute little to the hemispheric trend. In the Southern Hemisphere, both the South Pacific and the South Indian Ocean regions contribute substantially to the poleward trend.

Within the period 1982–2009, the latitude of LMI can be reanalysed using a globally homogenized record of intensity (ADT-HURSAT[3]). When this is done, the annual-mean time series exhibit similar variability and trends (Fig. 1a, b, blue lines), although the ADT-HURSAT-based trend has a greater amplitude than the best-track-based trend in the Northern Hemisphere and a lesser amplitude than the best-track-based trend in the Southern Hemisphere (Table 1), where the trend is no longer significant with 95% confidence. However, when both hemispheres are considered together they depict a global migration away from the deep tropics, and the best-track and ADT-HURSAT data exhibit similar poleward trends of 115 and 118 km per decade, respectively (Table 1). In this global view, the trends in the best-track and ADT-HURSAT data are consistent and both are statistically significant.

As found with the best-track data, the ADT-HURSAT-based time series exhibit large differences in trend amplitudes and statistical power when separated by ocean basin (Table 1 and Extended Data Fig. 2). The western North Pacific is the largest contributor to the trend in the Northern Hemisphere, and the eastern North Pacific also contributes significantly, unlike the best-track data from that region. The equatorward trend in the North Indian Ocean best-track data is not found in the ADT-HURSAT data, which shows essentially no trend in that region (the lack of any poleward trend in the North Indian Ocean might be expected given the confines of the basin and the close proximity of land to the

[1]NOAA National Climatic Data Center, University of Wisconsin/CIMSS, 1225 West Dayton Street, Madison, Wisconsin 53706, USA. [2]Program in Atmospheres, Oceans, and Climate, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [3]NOAA Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey 08540, USA.

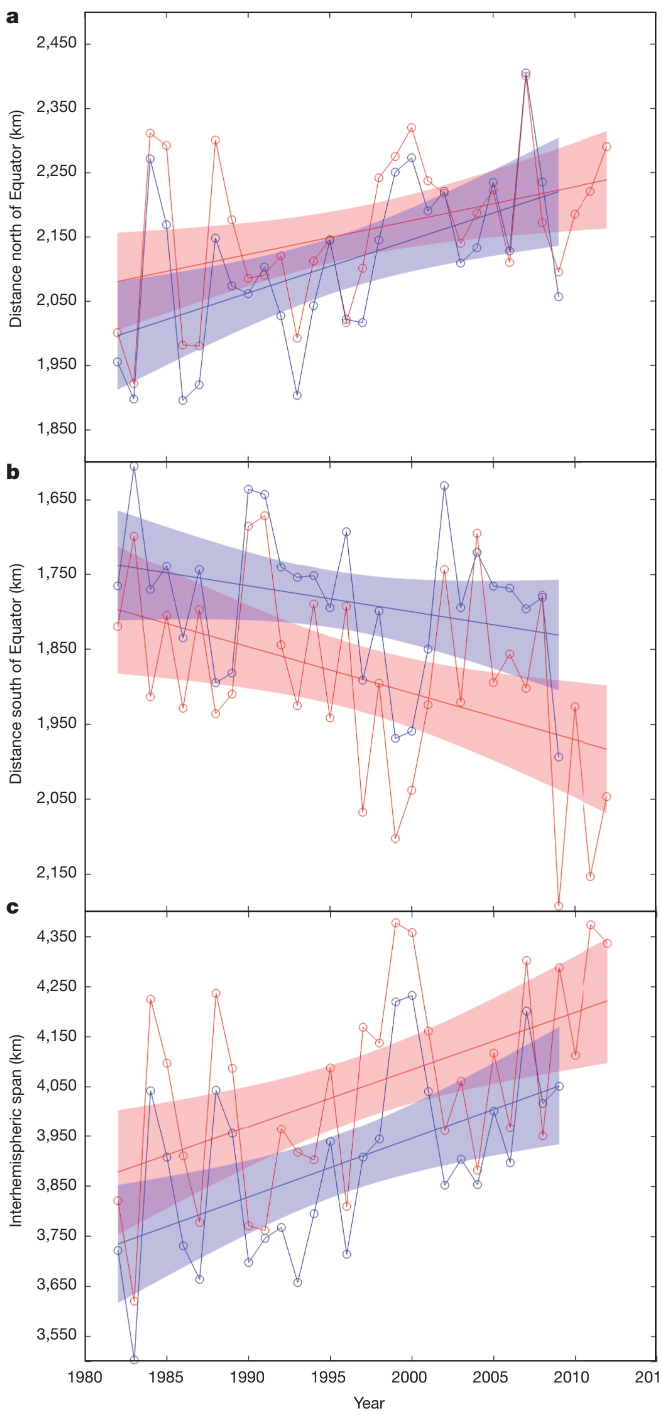


Figure 1 | Poleward migration of the latitude of LMI away from the tropics. **a**, **b**, Time series of annual-mean latitude of tropical cyclone LMI calculated from the best-track historical data (red) and the ADT-HURSAT reanalysis (blue) in the Northern (**a**) and Southern (**b**) hemispheres. **c**, The annual-mean difference between **a** and **b** shows the global migration of the latitude of LMI away from the tropics. Linear trend lines are shown with their 95% two-sided confidence intervals (shaded). Note that the *y* axis in **b** increases downwards.

north). In the North Atlantic, the best-track and ADT-HURSAT data sets both show essentially no trend. There are similar poleward trends in the best-track and ADT-HURSAT data from the South Pacific, but the ADT-HURSAT data in the South Indian Ocean exhibits a smaller, statistically insignificant trend.

Although regional differences are evident, the migration of the mean latitude of LMI away from the deep tropics is observed in both hemispheres, which indicates that this is a global phenomenon. The genesis and subsequent intensification period of tropical cyclones, which precedes LMI and controls when and where LMI occurs, is strongly modulated by the environment that the storms move through in this period. Known major factors controlling tropical cyclone evolution are the environmental vertical wind shear and the potential intensity[16–18]. Potential intensity describes the thermodynamically based maximum tropical cyclone intensity that the environment will support, all other factors being optimal. Vertical wind shear is one of the key factors that inhibit a storm from achieving this maximum. Greater shear and lesser potential intensity each inhibits genesis and intensification, and vice versa, and increased shear in the deep tropics, decreased shear at higher latitudes, or both, can thus be plausibly linked to a poleward migration of the latitude of LMI. Decreased potential intensity in the deep tropics, increased potential intensity at higher latitudes, or both, could be expected to result in a similar migration. Here we explore these environmental factors using three different atmospheric reanalysis products, NCEP/NCAR[12], ERA-Interim[13] and MERRA[14]. All three products exhibit broad regions of increased shear in the deep tropics and decreased shear in the subtropics (Fig. 2), which is consistent with the observed changes in the tropical cyclones. The changes in mean potential intensity are not as consistent among the different reanalysis products, particularly in the tropics, which is probably a result of spurious differences in upper tropospheric temperatures[19,20]. However, the meridional structure of potential intensity change is generally consistent in showing greater increases at higher latitudes, and the MERRA data, in particular, also show a broad reduction of potential intensity in the deep tropics.

The observed changes in shear and potential intensity provide evidence that the global migration of tropical cyclones away from the tropics is being modulated by systematic environmental changes. Shifts in tropical cyclone tracks in most regions have also been linked to phase changes in El Niño/Southern Oscillation[21–24] (ENSO), which can potentially contribute to the poleward trends in LMI identified here. To test this, we decrease the contribution of ENSO by regressing the LMI latitude time series onto an index of ENSO variability. When this is done (Fig. 3), the amplitude of the interhemispheric migration rates is found to decrease only slightly in both the best-track and the ADT-HURSAT data, and the statistical power of the trends in fact increases. This makes it unlikely that natural ENSO variability has a role in the observed multi-decadal poleward migration of LMI, although it plays a substantial part in its interannual variability.

The potential for contributions from natural variability occurring on decadal or longer timescales still exists, but quantifying this is difficult using relatively short observation records. We propose that there is a linkage between the poleward migration of LMI and the observed expansion of the tropics. The rate of expansion since 1979 varies considerably among existing studies[6], but the rate of LMI migration identified here falls well within this range. This potential linkage between tropical cyclones and the expansion of the tropics further heightens interest in establishing the forcing mechanisms of the expansion, which are at present uncertain but are generally thought to have anthropogenic contributions[6]. The expansion of the tropics, as measured by the meridional

Table 1 | Linear trends, by region, of annual-mean latitude of LMI

	NHEM	SHEM	NATL	WPAC	EPAC	NIO	SIO	SPAC	Global
Best track	+53 ± 43	+62 ± 48	+7 ± 98	+37 ± 55	+10 ± 32	−25 ± 78	+67 ± 55	+51 ± 68	+115 ± 70
ADT-HURSAT	+83 ± 50	+35 ± 44	−12 ± 126	+105 ± 71	+34 ± 30	+10 ± 106	+30 ± 52	+54 ± 79	+118 ± 70

Trends are deduced from the best-track and ADT-HURSAT data sets. The slope (kilometres per decade) and the 95% two-sided confidence bounds are shown. Positive slopes represent poleward migration. NHEM, Northern Hemisphere; SHEM, Southern Hemisphere; NATL, North Atlantic; WPAC, western Pacific; EPAC, eastern Pacific; NIO, North Indian Ocean; SIO, South Indian Ocean; SPAC, South Pacific.

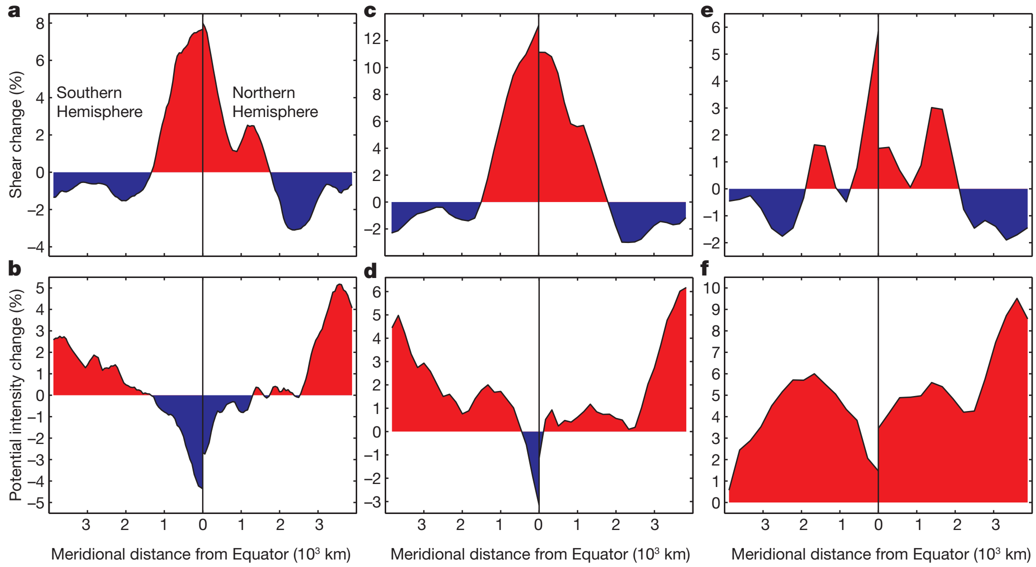


Figure 2 | Observed changes in the mean environment where tropical cyclones form and track. Percentage changes from 1980–1994 to 1995–2010 in mean vertical wind shear (**a**, **c**, **e**) and potential intensity (**b**, **d**, **f**). Annual means are taken over the peak tropical cyclone seasons in each hemisphere (August–October in the north and January–March in the south) from three different reanalysis products: MERRA (**a**, **b**), ERA-Interim (**c**, **d**) and NCEP/NCAR (**e**, **f**).

extent of the tropical Hadley circulation, exhibits a step change in the late 1990s[6]. Formal change-point analysis applied to the global time series of LMI latitude reveals a significant change point in 1996, providing further support for a linkage between the two independently observed phenomena.

Observed changes in vertical wind shear and potential intensity over the past 30 yr seem to have resulted in a poleward shift, in both hemispheres, of the regions most favourable for tropical cyclone development (Fig. 2), and an associated migration of tropical cyclone activity away from the tropics (Fig. 1). If these environmental changes continue, a concomitant continued poleward migration of the latitude where tropical cyclones achieve their LMI would have potentially profound consequences for life and property. Any related changes in positions where storms make landfall will have obvious effects on coastal residents and infrastructure. Increasing hazard exposure and mortality risk from tropical cyclones[25] may be compounded in coastal cities outside the tropics, while being offset at lower latitudes. Tropical cyclones also have an important role in maintaining regional water resources[26,27], and a poleward migration of storm tracks could threaten potable water supplies in some regions while increasing flooding events in others. Given these motivating factors, further study of the poleward migration of tropical cyclone LMI identified here, and its potential link to the observed expansion of the tropics, is warranted.

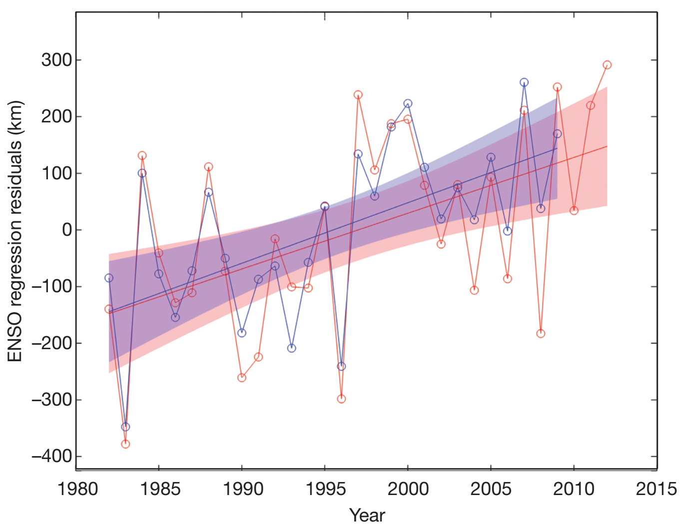


Figure 3 | Global trends of the latitude of LMI with ENSO variability reduced. Time series of the latitude of LMI calculated from the best-track historical data (red; trend, 99 $\pm$ 59 km per decade) and the ADT-HURSAT reanalysis (blue; trend, 107 $\pm$ 53 km per decade) with ENSO variability reduced. The values are calculated from residuals of the regression of LMI latitude onto an index of ENSO variability. Shading represents the 95% two-sided confidence interval of the trend.

METHODS SUMMARY

Best-track data were taken from the International Best Track Archive for Climate Stewardship (IBTrACS) v03r05 (ref. 28). Following ref. 3, when a storm has overlapping data from multiple sources, we used the source with the greatest reported LMI. The homogenized intensity data were taken from the Advanced Dvorak Technique Hurricane Satellite (ADT-HURSAT) data set[3]. Vertical wind shear and potential intensity were calculated over water in the region spanning latitudes 35° S to 35° N. In the Southern Hemisphere, the longitude was confined to 30°–240° E, which excludes the region where storms are not observed to form or track. The wind shear is estimated as the magnitude of the vector difference of the respective horizontal wind velocities at the 250- and 850-hPa pressure levels. Potential intensity was calculated following ref. 29. ENSO variability was decreased in the LMI latitude time series by regressing the series from the two hemispheres onto the Niño-3.4 index[30] averaged over the most active periods of tropical cyclone activity (August–October in the north and January–March in the south), and analysing the residuals. None of the time series studied in this paper exhibited autocorrelation after detrending, as determined with the Durbin–Watson test statistic, and no corrections were necessary when calculating the confidence intervals.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 21 October 2013; accepted 21 March 2014.

1. Landsea, C. W., Vecchi, G. A., Bengtsson, L. & Knutson, T. R. Impact of duration thresholds on Atlantic tropical cyclone counts. *J. Clim.* **23,** 2508–2519 (2010).
2. Vecchi, G. A. & Knutson, T. R. Estimating annual numbers of Atlantic hurricanes missing from the HURDAT database (1878–1965) using ship track density. *J. Clim.* **24,** 1736–1746 (2011).
3. Kossin, J. P., Olander, T. L. & Knapp, K. R. Trend analysis with a new global record of tropical cyclone intensity. *J. Clim.* **26,** 9960–9976 (2013).
4. Knutson, T. R. *et al.* Tropical cyclones and climate change. *Nature Geosci.* **3,** 157–163 (2010).

5. Seneviratne, S. I. *et al.* in *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation* (eds Field, C. B. *et al.*) 109–230 (Cambridge Univ. Press, 2012).
6. Lucas, C., Timbal, B. & Nguyen, H. The expanding tropics: a critical assessment of the observational and modeling studies. *WIREs. Clim. Change* **5,** 89–112 (2014).
7. Frank, W. M., &. Young, G. S. The interannual variability of tropical cyclones. *Mon. Weath. Rev.* **135,** 3587–3598 (2007).
8. Holland, G. & Bruyere, C. L. Recent intense hurricane response to global climate change. *Clim. Dyn.* **42,** 617–627 (2013).
9. Knapp, K. R. & Kruk, M. C. Quantifying interagency differences in tropical cyclone best track wind speed estimates. *Mon. Weath. Rev.* **138,** 1459–1473 (2010).
10. Emanuel, K. A. Increasing destructiveness of tropical cyclones over the past 30 years. *Nature* **436,** 686–688 (2005).
11. Bell, G. D. *et al.* Climate assessment for 1999. *Bull. Am. Meteorol. Soc.* **81,** s1–s50 (2000).
12. Kalnay, E. *et al.* The NCEP/NCAR 40-year reanalysis project. *Bull. Am. Meteorol. Soc.* **77,** 437–471 (1996).
13. Dee, D. P. *et al.* The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* **137,** 553–597 (2011).
14. Rienecker, M. *et al.* MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. *J. Clim.* **24,** 3624–3648 (2011).
15. Bindoff, N. L. *et al.* in *Climate Change 2013: The Physical Science Basis* (eds Stocker, T. F. *et al.*) 867–952 (Cambridge Univ. Press, 2013).
16. Knaff, J. A., Seseske, S. A., DeMaria, M. & Demuth, J. L. On the influences of vertical wind shear on symmetric tropical cyclone structure derived from AMSU. *Mon. Weath. Rev.* **132,** 2503–2510 (2004).
17. Zhang, F. & Tao, D. Effects of vertical wind shear on the predictability of tropical cyclones. *J. Atmos. Sci.* **70,** 975–983 (2013).
18. Emanuel, K. A. Thermodynamic control of hurricane intensity. *Nature* **401,** 665–669 (1999).
19. Vecchi, G. A., Fueglistaler, S., Held, I. M., Knutson, T. R. & Zhao, M. Impacts of atmospheric temperature trends on tropical cyclone activity. *J. Clim.* **26,** 3877–3891 (2013).
20. Emanuel, K., Solomon, S., Folini, D., Davis, S. & Cagnazzo, C. Influence of tropical tropopause layer cooling on Atlantic hurricane activity. *J. Clim.* **26,** 2288–2301 (2013).
21. Camargo, S. J., Robertson, A. W. & Gaffney, S. J. Smyth, P. & Ghil, M. Cluster analysis of typhoon tracks. Part II: large-scale circulation and ENSO. *J. Clim.* **20,** 3654–3676 (2007).
22. Camargo, S. J., Robertson, A. W., Barnston, A. G. & Ghil, M. Clustering of eastern North Pacific tropical cyclone tracks: ENSO and MJO effects. *Geochem. Geophys. Geosyst.* **9,** Q06V05 (2008).
23. Kossin, J. P., Camargo, S. J. & Sitkowski, M. Climate modulation of North Atlantic hurricane tracks. *J. Clim.* **23,** 3057–3076 (2010).
24. Ramsay, H. A., Camargo, S. J. & Kim, D. Cluster analysis of tropical cyclone tracks in the Southern Hemisphere. *Clim. Dyn.* **39,** 897–917 (2012).
25. Peduzzi, P. *et al.* Tropical cyclones: global trends in human exposure, vulnerability and risk. *Nature Clim. Change* **2,** 289–294 (2012).
26. Jiang, H. & Zipser, E. Contribution of tropical cyclones to the global precipitation from eight seasons of TRMM data: regional, seasonal, and interannual variations. *J. Clim.* **23,** 1526–1543 (2010).
27. Lam, H., Kok, M. H. & Shum, K. K. Y. Benefits from typhoons – the Hong Kong perspective. *Weather* **67,** 16–21 (2012).
28. Knapp, K. P., Kruk, M. C., Levinson, D. H., Diamond, H. J. & Neumann, C. J. The International Best Track Archive for Climate Stewardship (IBTrACS): unifying tropical cyclone data. *Bull. Am. Meteorol. Soc.* **91,** 363–376 (2010).
29. Bister, M. & Emanuel, K. A. Dissipative heating and hurricane intensity. *Meteorol. Atmos. Phys.* **65,** 233–240 (1998).
30. Barnston, A. G., Chelliah, M. & Goldenberg, S. B. Documentation of a highly ENSO-related SST region in the equatorial Pacific. *Atmosphere–Ocean* **35,** 367–383 (1997).

Author Contributions J.P.K. had the idea for, and designed, the study, and performed the analyses with input from K.A.E. and G.A.V. J.P.K., K.A.E. and G.A.V. provided data and participated in interpretation of the results and the writing of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.P.K. (james.kossin@noaa.gov).

METHODS

Best-track data were taken from the International Best Track Archive for Climate Stewardship (IBTrACS) v03r05 (ref. 28) and are available at http://www.ncdc.noaa.gov/ibtracs/. Following ref. 3, when a storm has overlapping data from multiple sources, we used the source with the greatest reported LMI. In storms that achieve their LMI more than once, the latitude of LMI is taken at the first occurrence. The homogenized intensity data were taken from the Advanced Dvorak Technique Hurricane Satellite (ADT-HURSAT) data set[3]. The data reflect the additional homogenization procedure addressing the discontinuity in satellite coverage that occurred in 1997. The global distribution of the ADT-HURSAT LMI is known to be spuriously leptokurtic[3], which is the likely cause of the consistent equatorward bias in the mean latitude of LMI when compared to the best-track data, but there is no expectation that this bias has any time dependence and it is not expected to affect the trends.

ENSO variability was removed from the LMI latitude time series by regressing the individual series from the Northern and Southern hemispheres onto the Niño-3.4 index[30] averaged over the most active periods of tropical cyclone activity (August–October in the north and January–March in the south), and analysing the residuals. The index is available at http://www.cpc.ncep.noaa.gov/data/indices/.

None of the time series explored in this paper exhibits autocorrelation after detrending, as determined with the Durbin–Watson test statistic, and no corrections were necessary when calculating the confidence intervals. In addition to linear trend analysis, the time series were explored for change points with models based on batch detection using both the Student *t* and Mann–Whitney statistics to test for significance at 95% confidence or greater using the 'cpm' package[31] in the software environment R.

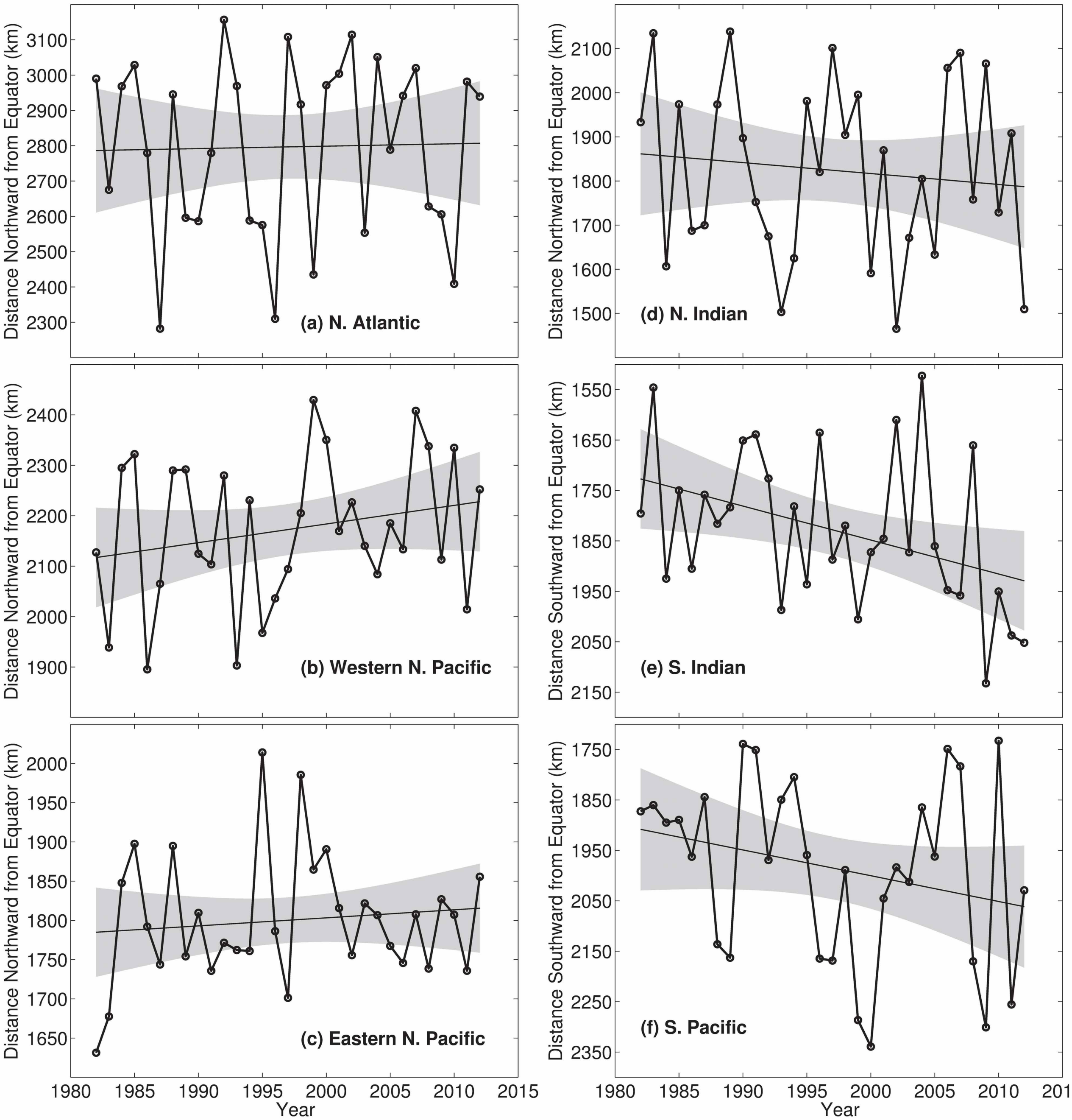
The global trends in the annual mean latitude of LMI are a result of both intrabasin and interbasin changes. The climatological mean latitude of LMI varies by basin (see, for example, Extended Data Fig. 1) such that, in addition to meridional shifts within each basin, changes in the relative annual frequency of storms from each basin can also contribute to the global trends in the latitude of LMI. To quantify this contribution, the LMI latitude of every storm was normalized by the respective basin-mean LMI latitude, and the analysis of Fig. 1c was repeated. When this was performed, the trend in the best-track data decreased from 115 ± 70 to 78 ± 66 km per decade and the trend in the ADT-HURSAT data decreased from 118 ± 70 to 92 ± 65 km per decade. Thus, both factors contribute, but the intrabasin poleward migration of LMI dominates the trends.

Monthly-mean vertical wind shear and potential intensity were calculated over water in the region spanning latitudes 35° S–35° N. In the Southern Hemisphere, the longitude was confined to 30°–240° E, which excludes the region where storms are not observed to form or track. Vertical wind shear was calculated as[32]

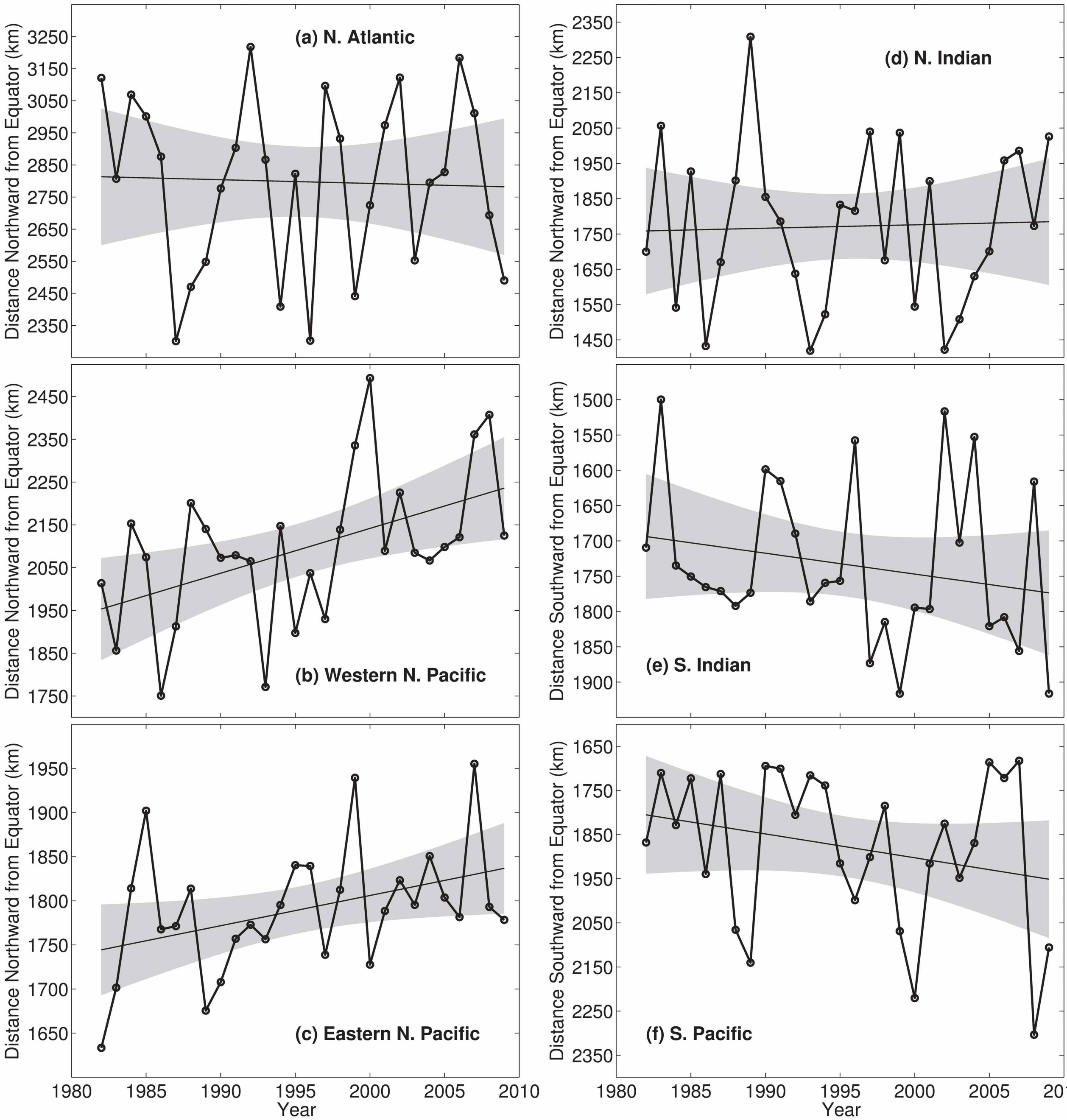
$$\text{shear} = \Big\{ (\bar{u}_{250} - \bar{u}_{850})^2 + (\bar{v}_{250} - \bar{v}_{850})^2 + \overline{u'^2_{250}} + \overline{v'^2_{250}} + \overline{u'^2_{850}} + \overline{v'^2_{850}} - 2\left(\overline{u'_{250}u'_{850}} + \overline{v'_{250}v'_{850}}\right) \Big\}^{1/2}$$

where u_{250}, u_{850}, v_{250} and v_{850} are the zonal (u) and meridional (v) winds at the 250- and 850-hPa pressure levels; u'_{250}, u'_{850}, v'_{250} and v'_{850} are departures of the corresponding daily means from their monthly means; and overbars represent monthly-mean quantities. Potential intensity was calculated following ref. 29. In the Northern Hemisphere the shear and potential intensity were averaged over August–October, and in the Southern Hemisphere they were averaged over January–March.

31. Ross, G. J. cpm: sequential parametric and nonparametric change detection. R package version 1.1. http://CRAN.R-project.org/package=cpm (2013).
32. Emanuel, K. Tropical cyclone activity downscaled from NOAA-CIRES reanalysis, 1908–1958. *J. Adv. Model. Earth Syst.* **2,** 1 (2010).



Extended Data Figure 1 | Time series of annual-mean latitude of LMI calculated from the best-track historical data from each ocean basin. The basins are those in the North Atlantic (**a**), the western North Pacific (**b**), the eastern North Pacific (**c**), the Northern Indian Ocean (**d**), the Southern Indian Ocean (**e**) and the South Pacific (**f**). Linear trend lines are shown with their 95% two-sided confidence intervals (shaded). Note that the *y* axes in **e** and **f** increase downwards.



Extended Data Figure 2 | Time series of annual-mean latitude of LMI calculated from the ADT-HURSAT data from each ocean basin. The basins are those in the North Atlantic (**a**), the western North Pacific (**b**), the eastern North Pacific (**c**), the Northern Indian Ocean (**d**), the Southern Indian Ocean (**e**) and the South Pacific (**f**). Linear trend lines are shown with their 95% two-sided confidence intervals (shaded). Note that the *y* axes in **e** and **f** increase downwards.

LETTER

doi:10.1038/nature13290

Anthropogenic electromagnetic noise disrupts magnetic compass orientation in a migratory bird

Svenja Engels[1,2]*, Nils-Lasse Schneider[1,2]*, Nele Lefeldt[1,2], Christine Maira Hein[1,2], Manuela Zapka[1,2], Andreas Michalik[1,2], Dana Elbers[1,2], Achim Kittel[3], P. J. Hore[4] & Henrik Mouritsen[1,2]

Electromagnetic noise is emitted everywhere humans use electronic devices. For decades, it has been hotly debated whether man-made electric and magnetic fields affect biological processes, including human health[1–5]. So far, no putative effect of anthropogenic electromagnetic noise at intensities below the guidelines adopted by the World Health Organization[1,2] has withstood the test of independent replication under truly blinded experimental conditions. No effect has therefore been widely accepted as scientifically proven[1–6]. Here we show that migratory birds are unable to use their magnetic compass in the presence of urban electromagnetic noise. When European robins, *Erithacus rubecula*, were exposed to the background electromagnetic noise present in unscreened wooden huts at the University of Oldenburg campus, they could not orient using their magnetic compass. Their magnetic orientation capabilities reappeared in electrically grounded, aluminium-screened huts, which attenuated electromagnetic noise in the frequency range from 50 kHz to 5 MHz by approximately two orders of magnitude. When the grounding was removed or when broadband electromagnetic noise was deliberately generated inside the screened and grounded huts, the birds again lost their magnetic orientation capabilities. The disruptive effect of radiofrequency electromagnetic fields is not confined to a narrow frequency band and birds tested far from sources of electromagnetic noise required no screening to orient with their magnetic compass. These fully double-blinded tests document a reproducible effect of anthropogenic electromagnetic noise on the behaviour of an intact vertebrate.

For more than 50 years, it has been known that night-migratory songbirds can use the Earth's magnetic field to orient spontaneously in their migratory direction when placed in an orientation cage at night in spring and autumn[7,8]. This basic experiment has been independently replicated many times in various locations[9]. We were therefore puzzled to find that night-migratory songbirds tested between autumn 2004 and autumn 2006 in wooden huts on the University of Oldenburg campus (53.1507° N, 8.1648° E) seemed unable to orient in the appropriate migratory direction. Typical data for European robins are shown in Fig. 1a.

Noting that Ritz *et al.*[10,11] had reported the sensitivity of European robins to radiofrequency magnetic fields, in the winter of 2006/2007 we decided to reduce the electromagnetic noise in our test huts by screening them with electrically connected and grounded aluminium plates (Extended Data Fig. 1). The screening left static magnetic fields such as the Earth's completely unaffected, but attenuated the electromagnetic noise inside the huts in the frequency range from about 50 kHz to at least 20 MHz by about two orders of magnitude (Fig. 1c, d and Methods). The effect on the birds' orientation capabilities was profound: with the aluminium screens in place, the birds oriented in their normal migratory direction the following spring (2007; Fig. 1b) and in subsequent years (data in references 12–15). When the horizontal component of the static magnetic field was rotated 120° anticlockwise or when the vertical component was inverted, the birds changed their orientation as expected[12–15].

These observations suggested that, by chance, we could have discovered a biological system that is sensitive to man-made electromagnetic noise in the range up to 5 MHz with intensities well below the guidelines for human exposure proposed by the International Commission on Non-Ionizing Radiation Protection (ICNIRP) and adopted by the World Health Organization[1,2].

Any report of an effect of low-frequency electromagnetic fields on a biological system should be subjected to particular scrutiny for at least

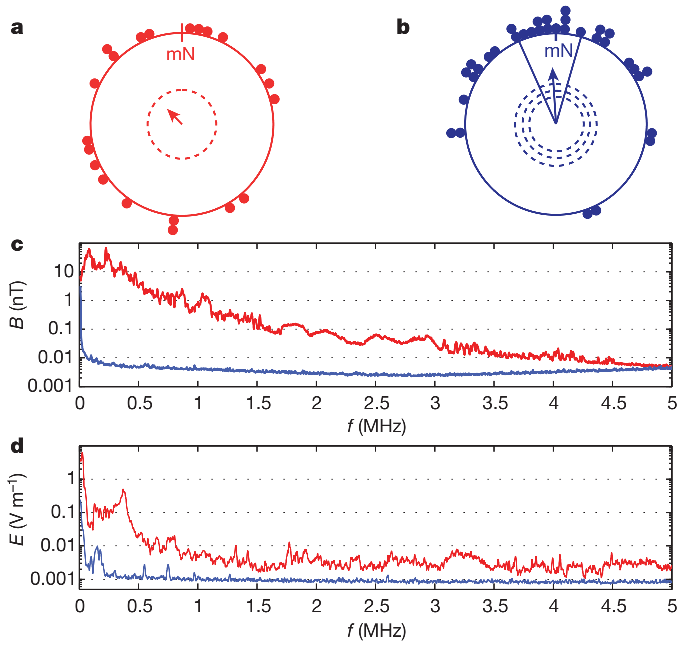


Figure 1 | Magnetic compass orientation of migratory European robins tested at the University of Oldenburg requires aluminium screening. In unscreened wooden huts, European robins were disoriented (**a**, spring 2005, $n = 21$, mean direction 316°, mean vector length $r = 0.19$, $P = 0.48$ (Rayleigh test)), but after installing grounded aluminium screens, the birds oriented highly significantly towards North in spring (**b**, spring 2007, $n = 34$, mean direction 356° ± 20° (95% confidence interval), $r = 0.59$, $P < 0.001$). **c**, **d**, Anthropogenic electromagnetic noise in the huts before (red) and after (blue) installation of screens. Traces **c** and **d** show the magnetic (B) and electric (E) components of the measured electromagnetic fields, respectively, as a function of frequency (f). In **a**, **b**, each dot indicates the mean orientation of all the tests of one individual bird in the given condition. The dots are colour-coded as in **c**, **d**. The arrows show group mean vectors flanked by their 95% confidence interval limits (solid lines). The dashed circles indicate the minimum length of the group mean vector needed for significance according to the Rayleigh test (inner circle, $P = 0.05$; middle, $P = 0.01$; outer, $P = 0.001$). mN, magnetic North.

[1]Institut für Biologie und Umweltwissenschaften, Universität Oldenburg, D-26111 Oldenburg, Germany. [2]Research Centre for Neurosensory Sciences, University of Oldenburg, D-26111 Oldenburg, Germany. [3]Institute of Physics, University of Oldenburg, D-26111 Oldenburg, Germany. [4]Department of Chemistry, University of Oxford, Physical and Theoretical Chemistry Laboratory, Oxford OX1 3QZ, UK.
*These authors contributed equally to this work.

three reasons. First, such claims in the past have often proved difficult to reproduce[1–6]. Second, animal studies are commonly used to evaluate human health risks and have contributed to guidelines for human exposures[1–4]. Third, "seemingly implausible effects require stronger proof"[16].

Therefore, we systematically conducted a large number of double-blind experiments over the last 7 years to test whether the restored orientation inside the aluminium-screened buildings was really attributable to the reduced exposure to anthropogenic electromagnetic noise. To ensure that our results were reliable, different generations of students independently replicated several key measurements. We also consulted with leading experts to ensure that we very carefully measured the electromagnetic fields experienced by the birds in each of the experimental conditions described below. Electromagnetic fields have magnetic and electric components, and, especially in the so-called 'near-field' (within a few wavelengths of the source), they must be measured separately.

First, we measured that the aluminium shielding lost its ability to screen anthropogenic electromagnetic noise when the grounding was disconnected (Fig. 2e, f). We therefore performed a series of experiments in which we tested a group of birds alternately in two different, aluminium-screened, wooden huts; one grounded and one left ungrounded. The experimenters were unaware which hut was which. The results were striking: on the days when the birds were tested in a grounded hut, they oriented in their mean northerly migratory direction as expected in spring (Fig. 2a, c). By contrast, the same birds were randomly oriented on the days when they were tested in an ungrounded hut (Fig. 2b, d). Thus, we could control the orientation of the birds inside the huts by connecting or disconnecting the grounding of the aluminium screens (Fig. 2).

Second, we assessed whether the electromagnetic noise was directly responsible for the disorientation. The birds were tested in the grounded aluminium-screened huts in which they normally orient very well (Figs 1b, 2a, c and data in references 12–15). The birds became disoriented (Fig. 3a) when we introduced broadband electromagnetic noise ranging from 2 kHz up to ca. 9 MHz (Fig. 3d, e and Extended Data Fig. 2) into the huts at magnetic field intensities similar to those measured for the background anthropogenic noise (Fig. 1c). To make sure that the observed effect was not simply due to the presence of the signal generator and associated electronics, we repeated these tests under identical conditions but with the output of the signal generator reduced to the lowest possible amplitude (Fig. 3d, e and Extended Data Fig. 2). In this condition, the birds oriented in their migratory direction in spring (Fig. 3b) and reoriented appropriately when the static magnetic field was rotated 120° anticlockwise (Fig. 3c). Thus, the disorientation appears to be caused by the electromagnetic noise, and not by the mere presence of the electronics.

Third, we assessed whether the effects are limited to a specific part of the radiofrequency spectrum. To answer this question, we tested European robins inside the grounded, aluminium-screened huts and in the presence of deliberately introduced broadband electromagnetic noise either in the frequency range from ca. 20 kHz to 450 kHz or from ca. 600 kHz to 3 MHz (Fig. 4f, g and Extended Data Fig. 2). As a control, we tested the same birds exposed to very-low-amplitude broadband noise ranging from ca. 2 kHz to 9 MHz (Figs 3d, e, 4f, g and Extended Data Fig. 2) in which we had observed that the birds could orient (Fig. 3b, c). As expected, the control birds again oriented appropriately (Fig. 4d, e). By contrast, broadband electromagnetic noise in both of the above non-overlapping frequency bands prevented the birds from orienting using their magnetic compass (Fig. 4a–c). Thus, the effects are not limited to one specific frequency or to one part of the radiofrequency spectrum.

The peak magnetic field intensity of the anthropogenic electromagnetic noise at any single frequency measured on typical days around the University of Oldenburg is on the order of 0.1–50 nT. The total time-dependent magnetic field, summed over the frequency range 10 kHz–5 MHz, is much stronger (on the order of at most 200–1,100 nT, see

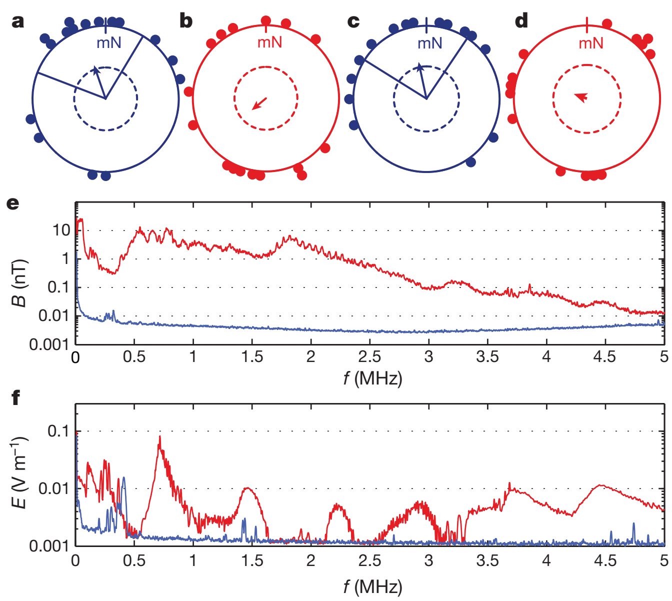


Figure 2 | Connecting and disconnecting the grounding of the screens turns on and off the birds' magnetic compass orientation capabilities. When the screens were grounded, European robins oriented significantly in their migratory direction (**a**, spring 2008, $n = 16$, mean direction 341° ± 40°, $r = 0.45$, $P = 0.04$), whereas they were randomly oriented when the grounding was disconnected (**b**, spring 2008, $n = 16$, mean direction 230°, $r = 0.22$, $P = 0.47$). In another set of identical tests, this pattern repeated itself (**c**, grounded screens, spring 2008, $n = 15$, mean direction 348° ± 41°, $r = 0.48$, $P = 0.03$; **d**, grounding disconnected, spring 2008, $n = 14$, mean direction 290°, $r = 0.12$, $P = 0.82$). **e**, **f**, Magnetic and electric field intensities, respectively, as a function of frequency when the screens were grounded (blue) or ungrounded (red).

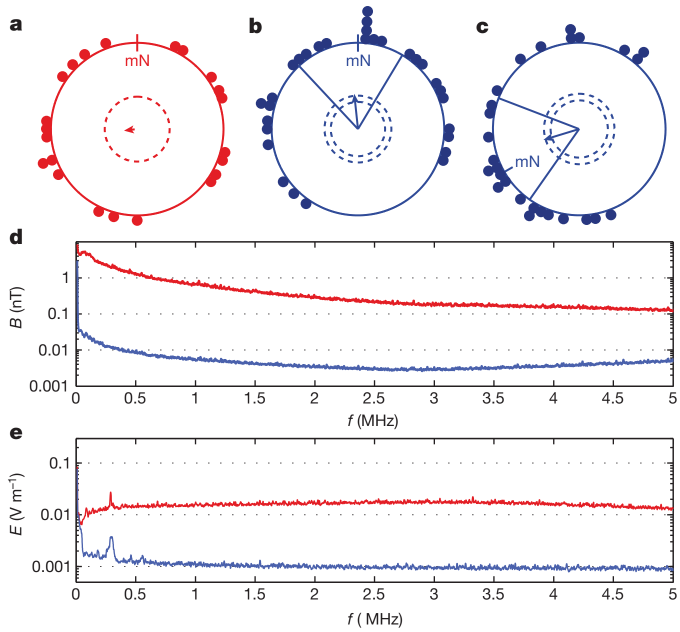


Figure 3 | Artificially produced broadband electromagnetic noise disrupts the magnetic compass orientation of birds tested inside the grounded aluminium-screened huts. Broadband, noise-modulated, electromagnetic fields between 2 kHz and 5 MHz (red traces in **d**, **e** and Extended Data Fig. 2) added inside the grounded screens resulted in disorientation of the birds (**a**, autumn 2010, $n = 22$, mean direction 278°, $r = 0.07$, $P = 0.91$). When the same equipment sent out the weakest possible broadband electromagnetic noise (blue traces in **d**, **e** and Extended Data Fig. 2), the birds oriented significantly towards North (**b**, spring 2011, $n = 30$, mean direction 354° ± 38°, $r = 0.39$, $P = 0.009$) and turned their orientation appropriately when the static magnetic field was rotated −120° (**c**, spring 2011, mN at 240°, $n = 27$, mean direction 253° ± 38°, $r = 0.41$, $P = 0.008$). **d**, Magnetic field intensity. **e**, Electric field intensity.

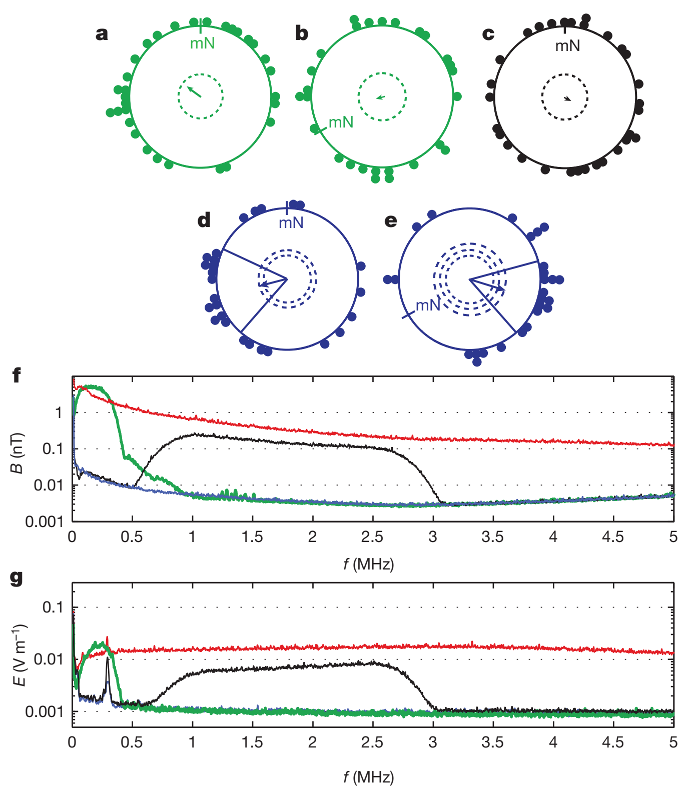


Figure 4 | The disruptive effect of broadband electromagnetic noise on magnetic compass orientation is not limited to a single narrow frequency range. Addition of broadband, noise-modulated, electromagnetic fields between ca. 20 kHz and 450 kHz (green traces in **f**, **g**) inside the grounded screens resulted in disorientation of the birds in the normal field (**a**, autumn 2011, $n = 31$, mean direction 306°, $r = 0.24$, $P = 0.17$) and in a field turned −120° horizontally (**b**, autumn 2011, $n = 27$, mean direction 235°, $r = 0.03$, $P = 0.96$). Broadband fields between ca. 600 kHz and 3 MHz (black traces in **f**, **g**) also disoriented the birds (**c**, autumn 2011, $n = 30$, mean direction 108°, $r = 0.11$, $P = 0.70$). When the same equipment sent out the weakest possible broadband electromagnetic noise (blue traces in **f**, **g**), the birds showed appropriately directed magnetic compass orientation (**d**, autumn 2011, $n = 27$, mean direction 258° ± 37°, $r = 0.42$, $P = 0.008$), and responded to a −120° horizontal rotation of the static field (**e**, autumn 2011, $n = 26$, mean direction 107° ± 32°, $r = 0.51$, $P < 0.001$). For comparison, the red traces in **f**, **g** show the intensity of the strong 2 kHz–9 MHz broadband noise used for the experiments presented in Fig. 3. **f**, Magnetic field intensity. **g**, Electric field intensity.

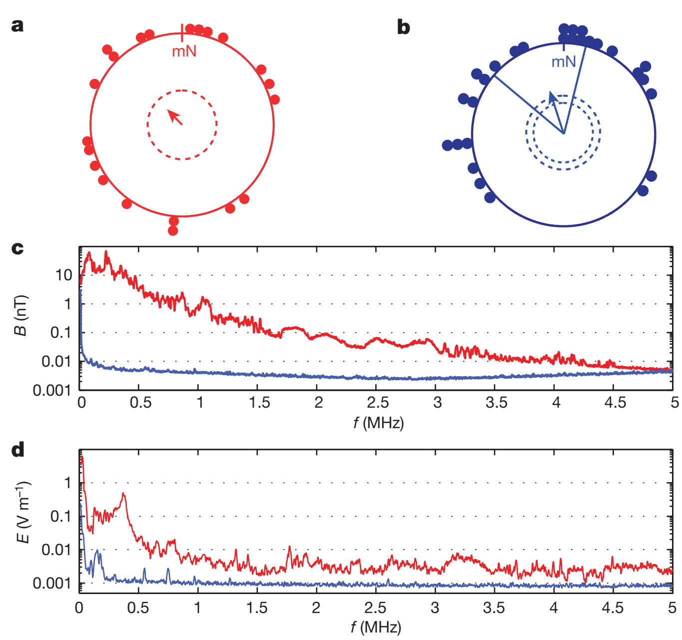


Figure 5 | In a rural location, European robins show magnetic compass orientation without screening. **a**, Orientation at the University campus (same data as in Fig. 1a). **b**, Orientation at a rural location (spring 2011, $n = 28$, mean direction 342° ± 32°, $r = 0.47$, $P = 0.002$) where the anthropogenic electromagnetic noise was much weaker (blue traces in **c**, **d**) than at the University (red traces in **c**, **d**). **c**, Magnetic field intensity. **d**, Electric field intensity.

Extended Data Table 1), but still much weaker than the Earth's magnetic field (ca. 49,000 nT in Oldenburg). Ritz *et al.*[11] reported that the magnetic orientation capabilities of European robins in Frankfurt were disabled by highly directional, monochromatic radiofrequency fields with magnetic field intensities of 15 nT or more, but not at 5 nT under otherwise identical conditions. Their birds were only disoriented at magnetic intensities below ca. 100 nT when the radiofrequency matched the electron Larmor frequency (1.315 MHz in Frankfurt; 1.363 MHz in Oldenburg), that is, the resonance frequency of the spin of a free electron interacting with the Earth's magnetic field. Electromagnetic fields similar to those used by Ritz *et al.*[10,11] never occur in natural or urban environments. The anthropogenic electromagnetic noise birds and other living beings experience is not monochromatic, nor is it spatially or temporally coherent (Fig. 1c, d). It has rapidly varying phases and directions and many different frequencies are present simultaneously. The electromagnetic noise we investigated is therefore fundamentally different from the conditions used previously[11]. Furthermore, our birds were never exposed to magnetic fields stronger than 1 nT at 1.315 MHz or 1.363 MHz (Figs 1–5), and two non-overlapping frequency ranges interfere with the birds' ability to use their magnetic compass (Fig. 4). Thus, the disruptive effect on orientation is not limited to a specific resonance frequency. It is caused by electromagnetic fields covering a much broader frequency range and at a much lower intensity (ca. 1 nT at any single frequency) than suggested previously[10,11]. Most importantly, broadband anthropogenic electromagnetic noise omnipresent in industrialized environments can lead to disorientation. These results have several important implications.

First, the present results could have significant consequences for migratory bird conservation. Magnetic compass information is sensed by night-migratory songbirds on the ground and in free flight[17,18], which mostly takes place at altitudes below 1,000 m (ref. 19). So, if anthropogenic electromagnetic fields prevent migratory songbirds from using their magnetic compass, their chances of surviving the migratory journey might be significantly reduced, in particular during periods of overcast weather when sun and star compass information is unavailable. Night-migratory songbird populations are declining rapidly[20], and anthropogenic electromagnetic noise could be a previously neglected contributory factor. Nevertheless, billions of migratory birds do find their way every year. It is therefore pertinent to ask, how localized is the disorienting effect of man-made electromagnetic noise?

We therefore compared the orientation of our robins in the unscreened huts at the University site (Figs 1a and 5a) with their orientation in an unscreened wooden shelter located ca. 7.5 km from the University and ca. 1 km outside the Oldenburg city limit, where the anthropogenic electromagnetic noise was much weaker (Fig. 5c, d) and similar in intensity to the electromagnetic noise remaining inside the grounded aluminium-screened huts (Fig. 1c, d, blue trace). In the rural setting, the birds could orient using their magnetic compass in the absence of screening (Fig. 5b). Thus, the disruptive effect of anthropogenic electromagnetic noise on the birds' orientation capabilities appears to be restricted to urban locations where there is typically a high usage of electronic devices. Therefore, the effect on wild birds is probably also quite localized.

Second, the results presented here are likely to provide key insights into the mechanism either of the magnetic compass sense[21–29] or of some important process that interferes with the birds' orientation behaviour. The biophysical mechanism that would allow such extraordinarily weak,

broadband electromagnetic noise to affect a biological system is far from clear. The energies involved are tiny compared to the thermal energy, $k_B T$, but the effects might be explained if hyperfine interactions in light-induced radical pairs[12,21–27] or large clusters of iron-containing particles are involved[28,29]. It would be truly remarkable if electromagnetic noise at the intensities studied here could be shown to disrupt the operation of a radical pair sensor by modifying its quantum spin dynamics. To be sensitive to such exceedingly weak magnetic fields, the electron spin-decoherence would have to be orders of magnitude slower than is currently thought possible. This intriguing prospect has attracted the attention of quantum physicists eager to learn lessons from nature that might ultimately allow more efficient quantum computers to be designed and constructed[30]. Furthermore, we cannot rule out that the birds might be affected by the electric component of the electromagnetic noise, a possibility that has not been considered previously.

Last, but not least, using a double-blinded protocol we have documented a clear and reproducible effect on a biological system of anthropogenic electromagnetic fields much weaker than the current ICNIRP guidelines[1,2]: the reference levels for general public exposure to time-varying magnetic fields in the relevant frequency band are 6,000 nT at 150 kHz decreasing to 180 nT at 5 MHz (refs 1, 2). The disruptive effects we observe cannot be attributed to power lines (16.7 Hz or 50 Hz fields) or to mobile phone signals (GHz frequencies) or to any other fields with frequencies below 2 kHz or above 5 MHz because outside this range the electromagnetic noise was of similar intensity in all conditions (Fig. 4 and Extended Data Fig. 2). Electromagnetic noise in the frequency-band 2 kHz–5 MHz originates primarily from AM radio signals and from electronic equipment running in University buildings, businesses and private houses. The effects of these weak electromagnetic fields generated by everyday human activity, however, are striking: they disrupt the function of an entire sensory system in a higher vertebrate.

METHODS SUMMARY

Essential methodological information needed for a basic understanding of the text has been woven into the main text at the appropriate places. The Methods section contains detailed information on the test subjects, electromagnetic shielding, execution and analysis of behavioural experiments, production and measurement of static fields, generation of electromagnetic noise, measurement of time-dependent electromagnetic fields, and blinding procedures.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 28 January; accepted 28 March 2014.
Published online 7 May 2014.

1. International Commission for Non-Ionizing Radiation Protection. ICNIRP guidelines for limiting exposure to time-varying electric, magnetic, and electromagnetic fields (up to 300 GHz). *Health Phys.* **74,** 494–522 (1998).
2. International Commission for Non-Ionizing Radiation Protection. ICNIRP statement on the ''Guidelines for limiting exposure to time-varying electric, magnetic, and electromagnetic fields (up to 300 GHz)''. *Health Phys.* **97,** 257–258 (2009).
3. World Health Organization. Extremely Low Frequency Fields. Environmental Health Criteria Monograph no. 238, http://www.who.int/peh-emf/publications/elf_ehc/en/ (2007).
4. Health Protection Agency. Health Effects from Radiofrequency Electromagnetic Fields http://www.hpa.org.uk/webc/HPAwebFile/HPAweb_C/1317133827077 (2012).
5. The INTERPHONE Study Group. Brain tumour risk in relation to mobile telephone use: results of the INTERPHONE international case–control study. *Int. J. Epidemiol.* **39,** 675–694 (2010).
6. Johansen, C. *et al.* Cellular telephones and cancer—a nationwide cohort study in Denmark. *J. Natl. Cancer Inst.* **93,** 203–207 (2001).
7. Merkel, F. W. & Wiltschko, W. Magnetismus und Richtungsfinden zugunruhiger Rotkehlchen. *Vogelwarte* **23,** 71–77 (1965).
8. Wiltschko, W. & Wiltschko, R. Magnetic compass of European robins. *Science* **176,** 62–64 (1972).
9. Wiltschko, R. & Wiltschko, W. *Magnetic Orientation in Animals* (Springer, 1995).
10. Ritz, T., Thalau, P., Phillips, J. B., Wiltschko, R. & Wiltschko, W. Resonance effects indicate a radical-pair mechanism for avian magnetic compass. *Nature* **429,** 177–180 (2004).
11. Ritz, T. *et al.* Magnetic compass of birds is based on a molecule with optimal directional sensitivity. *Biophys. J.* **96,** 3451–3457 (2009).
12. Zapka, M. *et al.* Visual but not trigeminal mediation of magnetic compass information in a migratory bird. *Nature* **461,** 1274–1277 (2009).
13. Hein, C. M. *et al.* Night-migratory garden warblers can orient with their magnetic compass using the left, the right or both eyes. *J. R. Soc. Interface* **7,** 227–233 (2010).
14. Hein, C. M., Engels, S., Kishkinev, D. & Mouritsen, H. Robins have a magnetic compass in both eyes. *Nature* **471,** E11 (2011).
15. Engels, S., Hein, C. M., Lefeldt, N., Prior, H. & Mouritsen, H. Night-migratory songbirds possess a magnetic compass in both eyes. *PLoS ONE* **7,** e43271 (2012).
16. Swanson, J. & Kheifets, L. Biophysical mechanisms: a component in the weight of evidence for health effects of power-frequency electric and magnetic fields. *Radiat. Res.* **165,** 470–478 (2006).
17. Cochran, W. W., Mouritsen, H. & Wikelski, M. Migrating songbirds recalibrate their magnetic compass daily from twilight cues. *Science* **304,** 405–408 (2004).
18. Chernetsov, N., Kishkinev, D., Kosarev, V. & Bolshakov, C. V. Not all songbirds calibrate their magnetic compass from twilight cues: a telemetry study. *J. Exp. Biol.* **214,** 2540–2543 (2011).
19. Bruderer, B. The study of bird migration by radar part 2: major achievements. *Naturwiss* **84,** 45–54 (1997).
20. Sanderson, F. J., Donald, P. F., Pain, D. J., Burfield, I. J. & Van Bommel, F. P. Long-term population declines in afro-palearctic migrant birds. *Biol. Conserv.* **131,** 93–105 (2006).
21. Schulten, K., Swenberg, C. E. & Weller, A. A biomagnetic sensory mechanism based on magnetic field modulated coherent electron spin motion. *Z. Phys. Chem.* **111,** 1–5 (1978).
22. Ritz, T., Adem, S. & Schulten, K. A model for photoreceptor-based magnetoreception in birds. *Biophys. J.* **78,** 707–718 (2000).
23. Maeda, K. *et al.* Chemical compass model of avian magnetoreception. *Nature* **453,** 387–390 (2008).
24. Rodgers, C. T. & Hore, P. J. Chemical magnetoreception in birds: the radical pair mechanism. *Proc. Natl Acad. Sci. USA* **106,** 353–360 (2009).
25. Liedvogel, M. & Mouritsen, H. Cryptochromes—a potential magnetoreceptor: what do we know and what do we want to know? *J. R. Soc. Interface* **7,** 147–162 (2010).
26. Mouritsen, H. & Hore, P. J. The magnetic retina: light-dependent and trigeminal magnetoreception in migratory birds. *Curr. Opin. Neurobiol.* **22,** 343–352 (2012).
27. Solov'yov, I. A., Domratcheva, T. & Schulten, K. Separation of photo-induced radical pair in cryptochrome to a functionally critical distance. *Sci. Rep.* **4,** 3845 (2014).
28. Kirschvink, J. L., Winklhofer, M. & Walker, M. M. Biophysics of magnetic orientation: strengthening the interface between theory and experimental design. *J. R. Soc. Interface* 7 Suppl. 2, S179–S191 (2010).
29. Wu, L. Q. & Dickman, J. D. Neural correlates of a magnetic sense. *Science* **336,** 1054–1057 (2012).
30. Gauger, E. M., Rieper, E., Morton, J. J. L., Benjamin, S. C. & Vedral, V. Sustained quantum coherence and entanglement in the avian compass. *Phys. Rev. Lett.* **106,** 040503 (2011).

Acknowledgements We thank M. Wuschek, Rohde & Schwarz, Bundesnetzagentur, and ETS Lindgren for help with measuring the electromagnetic fields, the workshops at the University of Oldenburg, especially T. Geiger, for building equipment, etc, and a large number of Bachelors, Masters and PhD students for help in conducting the experiments. We are grateful to the following for financial support: Defense Advanced Research Projects Agency (QuBE: N66001-10-1-4061 to P.J.H. and H.M.), VW-Stiftung (Lichtenberg professorship to H.M.), DFG (FOR 701 and MO 1408/2-2 to H.M.), Heinz Neumüller Stiftung (to C.M.H. and S.E.), BMBF (to H.M.), the European Research Council (to P.J.H.) and the EMF Biological Research Trust (to P.J.H.).

Author Contributions S.E. and N.-L.S. contributed equally to this work and are listed alphabetically. H.M. and N.-L.S. designed the study. S.E., N.L., C.M.H., M.Z., A.M. and D.E. performed the experiments. S.E., N.L., C.M.H., M.Z. and H.M. analysed the data. A.K., P.J.H. and N.-L.S. provided physical insight needed to properly produce and measure the electromagnetic fields. N.-L.S. and S.E. were in charge of generating the electromagnetic noise. N.-L.S. measured the electromagnetic fields. H.M., P.J.H., N.-L.S. and S.E. wrote most of the paper. All authors read and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.M. (henrik.mouritsen@uni-oldenburg.de).

METHODS

Test subjects. In our study, we tested European robins caught on the campus of the University of Oldenburg, Germany. The birds were housed indoors in individual cages in a windowless room under a light regime simulating the local photoperiod. The tests were performed on the campus of the University of Oldenburg during the spring migratory seasons in 2005 (when we tested 22 birds), 2008 (18 birds) and 2011 (30 birds) and during the autumn migratory seasons in 2010 (24 birds) and 2011 (42 birds). The number of birds caught during the previous migratory seasons and the experimental facilities available for the specific experiment in the given season determined the choice of sample sizes. In addition to these experiments, which were performed specifically for the present study, tests were also conducted by various groups of students in spring 2007[12], spring 2008[12], autumn 2008[13] (tests with garden warblers, *Sylvia borin*), spring 2009[12], autumn 2009[14,15], autumn 2010[15], and spring 2011[15]. These additional experiments, done primarily for other studies that have already been published[12–15], included tests with control groups which repeatedly confirmed and extended the results presented in Fig. 1, namely that: (a) night-migratory songbirds orient properly using their magnetic compass in the grounded and screened huts in the unchanged geomagnetic field[12–15], and (b) they adjust their orientation appropriately when the horizontal component of the static field is rotated by -120° (refs 12–15). Furthermore, in two previous studies[12,15] we tested groups of European robins in the screened and grounded huts while exposing them to a static field the vertical component of which had been inverted, leaving the horizontal component still pointing to the North. In this field, the polarity of the field lines is unchanged and still points towards magnetic North, but the axis of the static field lines is the same as if the static field had been turned 180° horizontally. Since these robins flipped their orientation ca. 180° (refs 12,15), the birds in the grounded and screened huts were using their standard magnetic inclination compass[8,9]. All animal procedures were approved by the Animal Care and Use Committees of the Niedersächsisches Landesamt für Verbraucherschutz und Lebensmittelsicherheit (LAVES, Oldenburg, Germany).

Static magnetic fields. Static magnetic fields were produced with double-wrapped, three-dimensional Merritt four-coil systems[31] with an average coil dimension of 2 m. All experiments were performed within the central space of the coils where the magnetic field homogeneity was better than 99%. Before the beginning of each experiment, the ambient geomagnetic field was measured using a Flux-Gate magnetometer (FVM-400, Meda Inc.) in the centre and at the edges of the experimental volume within which the orientation funnels were placed. Birds were tested in two different static magnetic field conditions: in a magnetic field closely similar to the natural geomagnetic field in Oldenburg (normal magnetic field, NMF) and in a magnetic field of the same strength and inclination as the local geomagnetic field but rotated 120° anticlockwise in the horizontal plane (changed magnetic field, CMF). To produce the CMF condition, the appropriate currents ran through the two subsets of windings per axis of the three-axial, four-coil Merritt system in the same direction. In the NMF condition, the currents that were needed to produce the CMF condition ran through the two subsets of windings but in opposite directions so that no significant changes (that is < 10 nT) to the geomagnetic field were produced by the coils[31]. The actual fields experienced by the birds under the two magnetic field conditions were as follows (mean $\pm$ standard deviation): NMF condition, 48,900 nT $\pm$ 150 nT; inclination, $67.7^\circ \pm 0.6^\circ$; horizontal direction, $360^\circ \pm 0.1^\circ$. CMF condition: 49,000 nT $\pm$ 470 nT; inclination, $68.0^\circ \pm 1.1^\circ$; horizontal direction, $-120^\circ \pm 0.5^\circ$.

Electromagnetic shielding of experimental huts. Most of the experiments were performed inside wooden huts (Extended Data Fig. 1a) placed at the Wechloy (Natural Sciences) Campus of the University of Oldenburg (Extended Data Fig. 1d) in the city of Oldenburg (population ca. 160,000; Extended Data Fig. 1c). Some of the orientation experiments in spring 2011 took place in an unscreened wooden shelter, normally occupied by horses, located in fields ca. 7.5 km from the University and ca. 1 km outside the built-up part of the city of Oldenburg (Extended Data Fig. 1c). An earth barrier in the form of a highway bridge foundation was located between the testing location and the city of Oldenburg.

To attenuate time-dependent electromagnetic fields inside the wooden huts, the four walls (including the door) and the roof were covered with 1-mm-thick aluminium sheets, overlapping by at least 20 mm and bolted together with self-cutting screws every 5–10 cm (Extended Data Fig. 1b). We also tested whether the efficiency of the screens could be improved by adding aluminium sheets on the floor. No improvement was found, probably because negligible electromagnetic noise comes from below. Most of the experiments were therefore performed in huts screened on five sides in which the air-circulation was improved and the humidity variability reduced compared to shielding on six sides.

The aluminium walls of this five-sided Faraday cage were interconnected at all times. In the grounded conditions, this aluminium screening assembly was electrically connected at a single location to a single grounding rod with a depth of 8 m. In the ungrounded conditions, the grounding rod was manually disconnected from the aluminium screening assembly. Disconnection of the grounding removed the screening effect of the aluminium shields. In fact, the ungrounded aluminium screens acted as an antenna that slightly increased the magnetic field intensity at some frequencies inside the test chambers compared to the unscreened condition (compare Fig. 1c, d with Fig. 2e, f). The disconnection of the grounding during the critical grounding/ungrounding experiments (Fig. 2) was performed by a member of the laboratory who was not involved in the behavioural experiments, and the persons performing and evaluating the experimental results were not aware of the change in conditions until after the completion of the experiments.

All electronic devices were placed outside this cage, disconnected from their protected earths and grounded via the same grounding rod as the Faraday cage. This is necessary because the protected earth from the standard power outlet would act as an antenna and introduce electromagnetic noise into the system. When properly grounded, the shielding attenuated the time-dependent magnetic fields with frequencies up to at least 20 MHz by approximately two orders of magnitude inside the testing chambers. The screening efficiency was estimated by generating electromagnetic noise just outside the chambers while measuring the electromagnetic noise arriving within. The anthropogenic electromagnetic noise observed at the University of Oldenburg is dominated by frequencies below 5 MHz. Higher frequency contributions were mostly at or below the detection limit of our equipment and are therefore not shown in Figs 1–5.

Generation of electromagnetic noise. To produce electromagnetic noise, a passive loop antenna (ETS Lindgren, Model 6511, 20 Hz–5 MHz) was placed vertically under the centre of the central orientation funnel and aligned along the North–South axis (48 cm vertically from the centre of the loop to the central funnel).

Broadband electromagnetic noise in the range 2 kHz–9 MHz was produced by a signal generator (Hewlett Packard, 33120A, 15 MHz Arbitrary Waveform Generator) connected to the antenna using either the maximum output (10 V peak-to-peak (pp)) for the strong noise condition or the minimum output (50 mV pp) for the weak noise condition used as a control (the generated noise with the output set to 50 mV pp was weaker than the measurement limit except for the electric component below 500 kHz, see blue traces in Fig. 3d, e and Extended Data Fig. 2). An alternative to this control would have been to use the 'silent shorting' design suggested by Kirschvink *et al.*[28]. We experimented with this method, but even the shorted condition led to measurably increased electromagnetic fields inside the huts, which is why we chose the control described above.

The band-pass electromagnetic noise (20 kHz–450 kHz and 600 kHz–3 MHz) was produced using a vector signal generator (Rohde & Schwarz, SMBV 100A, 9 kHz–3.2 GHz) connected to the same passive loop antenna.

Measurements of time-dependent electromagnetic fields. The magnetic and electric components of the time-dependent electromagnetic fields were measured separately with different antennas connected to a signal analyser (Rohde & Schwarz, FSV 3 Signal and Spectrum Analyzer 10 Hz–3.6 GHz). All such measurements were performed at a similar time of day as the behavioural experiments, but not while the actual tests were running. This procedure was chosen because we wanted to exclude any possibility that the measurements or measuring equipment could influence in any way the electromagnetic noise fields present while the birds were being tested.

The magnetic component between 10 kHz and 5 MHz was measured with a calibrated passive loop antenna (ETS Lindgren, Model 6511, 20 Hz–5 MHz). The electric component between 10 kHz and 10 MHz was measured with a calibrated active biconical antenna (Schwarzbeck Mess-Electronik, EFS 9218, 9 kHz–300 MHz). The signal analyser was set to 'max hold' and a resolution bandwidth of 10 kHz. For each condition we measured the fields for a period of 40 min. The traces shown in Figs 1–5 are based on 5,000 measurement points between 10 kHz and 5 MHz.

For the low-frequency range (5 Hz–32 kHz), we used the EFA-300 system (Narda Safety Solutions). The magnetic component was measured using the calibrated EFA Magnetic Field Probe 100 cm^2 (EFA-300 system, Narda Safety Solutions). The electric component was measured with the calibrated Narda Electric Field Unit (EFA-300 system, Narda Safety Solutions). For each measurement, the antennas were connected to the EFA-300 hand-held signal analyser, and this signal analyser was also set to 'max hold' and the fields were measured for a period of 40 min (Extended Data Fig. 2).

It must be stressed that anthropogenic electromagnetic noise fields are always present but highly variable in their amplitude, phase and frequency spectrum. Two measurements of their intensity and frequency composition will never be identical. Consequently, the measurements shown in Figs 1c, d, 2e, f and 5c, d are representative examples of the noise measured at the approximate time of day when the experiments were performed.

The maximal total magnetic field intensity (more precisely the magnetic flux density, B) in the frequency range between 10 kHz and 5 MHz was calculated using the following equation:

$$B(\Delta f) = \frac{1}{N}\frac{\Delta f}{\Delta f_0}\sum_i B_i(f_i, \Delta f_0)$$

in which $B(\Delta f)$ denotes the total magnetic flux density in the bandwidth of interest, $\Delta f = 5$ MHz–10 kHz = 4,990 kHz, and $B_i(f_i, \Delta f_0)$ is the magnetic flux density at the N different frequency values f_i (every 1 kHz between 10 kHz and 5 MHz, that is, $N = 4{,}990$) for a resolution bandwidth Δf_0, which equals 10 kHz here. Expressed in words, $B(\Delta f)$ = (the sum of the magnetic field intensity values/no. of values) × (frequency range size/resolution bandwidth), in our case: (the sum of the magnetic field intensity values/4,990) × (4,990 kHz/10 kHz) for the total frequency range from 10 kHz to 5,000 kHz. Extended Data Table 1 lists these values for the different conditions tested.

Behavioural experiments. All birds were tested in so-called Emlen funnels[32] lined with scratch-sensitive paper[33], inside wooden huts (4 m × 4 m × ca. 3 m, Extended Data Fig. 1a), where no directional cue other than the geomagnetic field was available. In 2005, the experiments took place in these simple wooden huts. From 2007 onwards, the walls and ceilings of the huts were lined with aluminium shields as described above. All electronic equipment was placed outside the hut in a separate wooden annex inside an aluminium box and grounded to minimize the generation of electromagnetic noise by the equipment itself.

One hour (± 10 min) before the experiments started (half an hour before until half an hour after sunset), the birds were placed outdoors in wooden transport cages that allowed them to see parts of the evening sky. This gave the birds the possibility to calibrate their magnetic compass from twilight cues[17,34]. Immediately thereafter, the birds were placed in modified aluminium Emlen funnels (35 cm diameter, 15 cm high, walls 45° inclined[32]), which were coated with thermal paper[33] on which the birds left scratches as they moved. The top of each funnel was covered with a translucent Plexiglas lid that prevented the birds from seeing any landmarks in the hut. The overlap point of the paper was adjusted to one of the cardinal directions (N, S, E or W). This overlap point was changed randomly between huts and nights. This is important because the papers are always evaluated relative to the overlap point by researchers who do not know in which direction it was positioned. Even if someone would intentionally try to ignore the condition-blinding protocols (this is highly unlikely), this procedure adds a second level of blinding, and it becomes impossible for 'wishful thinking' to influence the results in any way, since the persons evaluating the papers cannot know which geographical direction is equivalent to a given direction on the paper. The location of the overlap point is only revealed and taken into consideration after the primary evaluation of the papers has taken place (for procedures see below).

The birds were tested for 1 h under dim white light conditions (2.1 mW m^{-2}) produced by incandescent bulbs (spectrum given in ref. 12). In each hut, nine birds were tested simultaneously. The birds were placed in a randomized funnel position each night and were put into the funnels from different directions, and we observed no systematic differences between the nine funnel positions or between the four huts. A second, and in a few instances a third, round of tests on a given night started 1.5 h (± 10 min) after the first or second round. In most cases, each bird was tested in a different hut in each round but under the same magnetic field condition (NMF or CMF) and if applicable under the same time-dependent electromagnetic noise condition. The results of the different tests can therefore be treated as independent. The mean direction of each bird in each condition was calculated by unit-vector addition of the individual mean directions from the typically 3–15 tests per bird per condition in which the bird was judged to be oriented.

If more than one condition was tested in a given season, the same experimental birds were tested in all conditions. The experimental condition experienced by a given bird was mostly interchanged every second day, and whenever possible, different conditions were run simultaneously in different huts, so that any putative daily variation, for instance induced by the weather[35], would be averaged out amongst the experimental groups.

In spring 2008, we decided to test the effect of the grounding of the shielding and performed experiments in two different huts. One of them was grounded (g) and the other was left ungrounded (u) without the experimenters knowing which one was which. The experimental condition for each bird alternated every other day; half the birds were tested in g-u-g-u conditions while the other half were u-g-u-g as follows: group 1, grounded on days 1 and 2, ungrounded on days 3 and 4, grounded on days 5 and 6, and ungrounded on days 7 and 8; group 2, ungrounded on days 1 and 2, grounded on days 3 and 4, ungrounded on days 5 and 6, and grounded on days 7 and 8. The data from these measurements are presented in Fig. 2 as follows: Fig. 2a: group 1: days 1, 2 and group 2: days 3, 4. Figure 2b: group 2: days 1, 2 and group 1: days 3, 4. Figure 2c: group 1: days 5, 6 and group 2: days 7, 8. Figure 2d: group 2: days 5, 6 and group 1: days 7, 8.

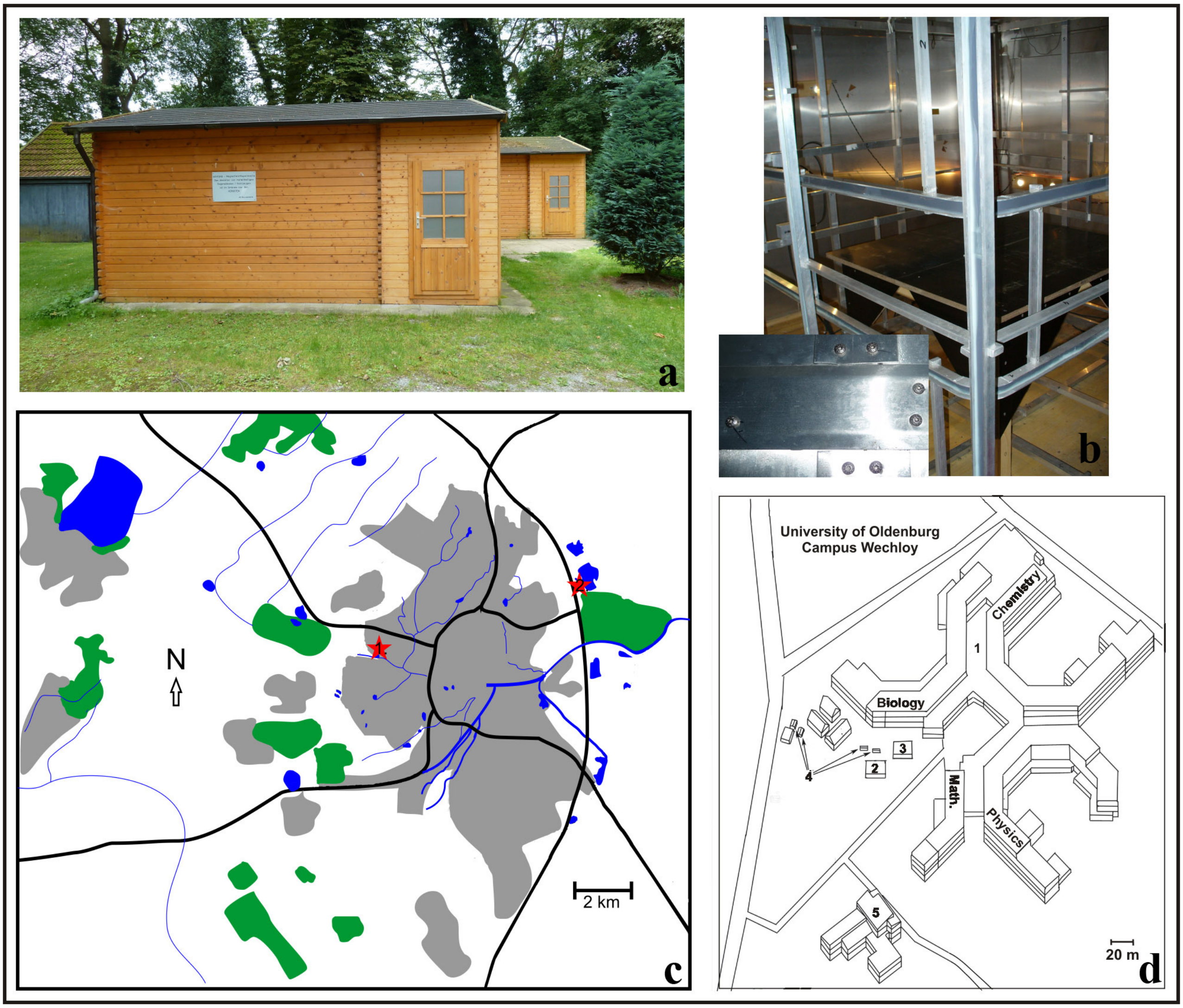
In 2010 and 2011, we performed experiments in which we added broadband electromagnetic noise (for details see above). The direction of the static magnetic field and electromagnetic noise conditions in a given hut were changed regularly; usually different conditions were tested concurrently in different huts on any given night.

At the rural location, 12 European robins were tested simultaneously in a wooden shelter located in a field (Extended Data Fig. 1c). Here, the birds were tested under natural magnetic conditions without a magnetic coil system. Other testing procedures were the same as in the huts on the University campus.

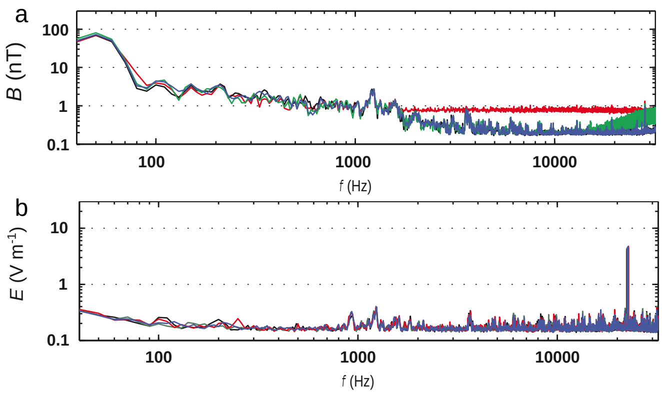
Before we started the experiments in any migratory season, we tested the birds in NMF and CMF conditions with no experimental manipulation for several nights to ensure that they were in migratory mood and to get a control direction.

Orientation data analysis. Independently, two researchers visually determined each bird's mean direction to the nearest 10° from the distribution of the scratches without knowing the direction of the overlap-point of the paper or the magnetic field conditions experienced by the bird. If one of the two researchers considered the scratches to be randomly distributed and the other did not, or if the two independently determined mean directions deviated by more than 30°, a third independent researcher was asked to determine the mean direction. If this third individual determined a mean direction similar to one of the first two, and if the individual with initially differing opinion also agreed with this direction, the mean of the two similar directions was recorded as the orientation result. If the three independent researchers could not agree on a mean direction, the bird's heading was defined as random and excluded from the analyses (7% of all tests). Birds with fewer than the pre-established lower limit of 100 scratches on the paper were considered inactive[15] and were also excluded from the analysis (40% of all tests). The observers performed this screening before they knew the direction of the overlap-point (see above). In this way we can be certain that the person making the decision on whether the bird left more or less than 100 scratches was not influenced by the bird's directional preferences. The average mean heading for each bird was calculated from all its oriented tests recorded under a given experimental condition. On the basis of these individual mean directions, group mean vectors were calculated by summing unit vectors in the mean directions of each individual bird and dividing by the total number of birds tested. The significance of the group mean vector was tested using the Rayleigh test[36].

31. Kirschvink, J. L. Uniform magnetic fields and double-wrapped coil systems: improved techniques for the design of bioelectromagnetic experiments. *Bioelectromagnetics* **13,** 401–411 (1992).
32. Emlen, S. T. & Emlen, J. T. A technique for recording migratory orientation of captive birds. *Auk* **83,** 361–367 (1966).
33. Mouritsen, H., Feenders, G., Hegemann, A. & Liedvogel, M. Thermal paper can replace typewriter correction paper in Emlen funnels. *J. Ornithol.* **150,** 713–715 (2009).
34. Muheim, R., Phillips, J. B. & Åkesson, S. Polarized light cues underlie compass calibration in migratory songbirds. *Science* **313,** 837–839 (2006).
35. Hein, C. M., Zapka, M. & Mouritsen, H. Weather significantly influences the migratory behaviour of night-migratory songbirds tested indoors in orientation cages. *J. Ornithol.* **152,** 27–35 (2011).
36. Batschelet, E. *Circular Statistics in Biology* (Academic, 1981).



Extended Data Figure 1 | Wooden huts and experimental locations. **a**, Photograph of one of the four identical wooden huts used for our experiments. **b**, Photograph from the inside of an experimental hut showing the aluminium screening, parts of the Merritt coil system, and the table on which the funnels were placed. The insert shows the self-cutting screws used to connect the aluminium plates. **c**, Simple map of the city of Oldenburg. Built-up areas are shown in grey and nature-protected areas in green. Black lines denote highways, blue denotes water. Red stars: '1' indicates the location of the University campus and '2' the rural location used for some of the tests. **d**, Map of the University of Oldenburg Wechloy Campus. 1, main University building housing the biology, chemistry, physics and mathematics institutes; 2, botanical greenhouse; 3, iron-free wooden building; 4, the locations of the four wooden huts used for our experiments; 5, 'Next Energy' building.



Extended Data Figure 2 | Electromagnetic noise measurements in the range from 40 Hz to 32 kHz. **a**, Magnetic field intensity (*B*). **b**, Electric field intensity (*E*). The colour coding of the traces corresponds to Fig. 4. Notice that the frequency-axis (*f*) is logarithmic.

Extended Data Table 1 | The accumulated time-dependent magnetic field intensity summed over all the frequencies in the spectra recorded for each behavioural test condition

	Oldenburg unshielded	**Oldenburg grounded shielding**	**Oldenburg ungrounded shielding**	**Rural location**	**10 kHz - 5 MHz strong**	**10 kHz - 5 MHz weak**	**20 - 450 kHz bandpass**	**600 kHz - 3 MHz bandpass**
	Fig. 1 and 5 red trace	Fig. 1 and 2 blue trace	Fig. 2 red trace	Fig. 5 blue trace	Fig. 3 and 4 red trace	Fig. 3 and 4 blue trace	Fig. 4 green trace	Fig. 4 black trace
***f* (kHz)**	**Accumulated field intensity (nT)**							
10 - 5000	1008.10	2.56	827.86	2.60	278.88	3.31	133.09	34.83
100 - 5000	714.17	1.98	705.88	1.74	229.56	2.35	101.11	33.98
20 - 450	855.63	0.38	125.59	0.31	119.30	0.69	128.27	0.71
600 - 3000	81.89	0.87	561.05	0.79	98.81	1.00	1.16	32.48

doi:10.1038/nature13178

Dynamics and associations of microbial community types across the human body

Tao Ding[1] & Patrick D. Schloss[1]

A primary goal of the Human Microbiome Project (HMP) was to provide a reference collection of 16S ribosomal RNA gene sequences collected from sites across the human body that would allow microbiologists to better associate changes in the microbiome with changes in health[1]. The HMP Consortium has reported the structure and function of the human microbiome in 300 healthy adults at 18 body sites from a single time point[2,3]. Using additional data collected over the course of 12–18 months, we used Dirichlet multinomial mixture models[4] to partition the data into community types for each body site and made three important observations. First, there were strong associations between whether individuals had been breastfed as an infant, their gender, and their level of education with their community types at several body sites. Second, although the specific taxonomic compositions of the oral and gut microbiomes were different, the community types observed at these sites were predictive of each other. Finally, over the course of the sampling period, the community types from sites within the oral cavity were the least stable, whereas those in the vagina and gut were the most stable. Our results demonstrate that even with the considerable intra- and interpersonal variation in the human microbiome, this variation can be partitioned into community types that are predictive of each other and are probably the result of life-history characteristics. Understanding the diversity of community types and the mechanisms that result in an individual having a particular type or changing types, will allow us to use their community types to assess disease risk and to personalize therapies.

Building on previous analysis of a healthy cohort of 300 individuals, we analysed a 16S rRNA gene sequence data set from the HMP Consortium[2,3]. The final data release for this cohort provided 16S rRNA gene sequence data and clinical metadata (Extended Data Table 1) from two time points for each of 300 healthy individuals and from a third time point for 100 of the individuals at 15 body sites for men and 18 for women[5]; the interval between samplings varied between 30 and 451 days (median = 224 days). A significant difficulty in analysing microbiome data has been the considerable intra- and interpersonal variation in the composition of the human microbiome[3,6,7]. A recently proposed approach for overcoming this difficulty within the gastrointestinal tract has been the concept of enterotypes, or more generically, stool community types[4,8,9]. In this approach samples are clustered into bins based on their taxonomic similarity. Specific enterotypes have been associated with the amount of protein, fat and carbohydrates in one's diet, obesity, inflammatory bowel disease, and Crohn's disease[4,9–11]. Others have found associations between specific vaginal community types and the sexually transmitted *Trichomonas vaginalis*, pH, and ethnicity[12–14] and associations between skin community types and psoriasis[15]. Using bacterial community structures collected from 18 body sites and up to three time points, we applied community typing analysis to understand better the factors that affect the structure of the microbiome and contribute to human health.

Concern has been expressed regarding whether community types reflect partitioning of an abundance gradient or the presence of clusters of relative abundance profiles[8,16]. Two general approaches have been developed to assign samples to community types: partitioning around the medoid (PAM) and Dirichlet multinomial mixture (DMM) models[4,8]. To compare these methods we first generated simulated communities where there were one or four community types. Analysis of the simulated communities indicated that the negative log model evidence metric used by the DMM-based approach was superior to the metrics used to assess clusters within the PAM-based approach (Supplementary Information). Next, we assigned the samples for each body site to community types using both methods. Calculation of the negative log model evidence demonstrated that the community types identified using DMM were superior to those identified using the PAM-based approach (Extended Data Table 2 and Extended Data Fig. 1). Thus, our analysis of simulated data and the HMP data suggests that the community types represent clusters of relative abundance profiles.

Using the DMM-based approach, we identified between two (anterior nares) and seven (tongue dorsum) community types per body site (see Source Data associated with Fig. 1 for community data and DMM fits). As an example, bacteria from stool samples fell into four distinct community types (Fig. 1a). We observed that 63 genera were needed to account for 90% of the difference between a model with a single community type and four community types (see Source Data associated with Fig. 1). Thus, it was not merely the most abundant bacterial population that differentiated the types as has been previously reported (for example, *Bacteroides*, *Prevotella*, or *Ruminococcus*)[8–10,17]; rather, community types were identified based on complex configurations of numerous taxa. In fact, this supports the findings of the original study; that is, the taxa that typify each enterotype represent networks of co-occurring bacterial populations[9]. Inspection of the five most important genera, which accounted for 54% of the difference in fit between four community types and one, indicated that each community type represented a cluster of relative abundance profiles (Fig. 1b). Community type A had the highest levels of *Bacteroides* but lacked *Prevotella* and Ruminococcaceae. Similar to community type A, community type C also lacked *Prevotella*, but had a lower relative abundance of *Bacteroides* and had higher levels of *Alistipes*, *Faecalibacterium* and Ruminococcaceae. Community type D had fewer *Bacteroides* than community types A and C, but had higher levels of *Prevotella*. Community type B had the fewest *Bacteroides* and was dominated by a variety of populations affiliated within the Firmicutes. Furthermore, the diversity of the samples assigned to each of the community types indicated that type A had a significantly lower diversity than the other three types ($P < 0.001$). Community types A, C and D resembled the previously identified *Bacteroides*, *Ruminoccocus* and *Prevotella* enterotypes, respectively[9,10,17]. Analysis of the other body sites yielded analogous patterns.

Using the responses that subjects gave to an extensive survey (summarized in Extended Data Table 1), we identified demographic and life-history characteristics that could be correlated with different community types at each body site. Of the numerous characteristics tested, we observed significant associations between community types and whether the subject was ever breastfed, their gender, and their education level (see Source Data associated with Fig. 1). Whether an individual was ever breastfed was strongly associated with their stool community type ($P = 1 \times 10^{-4}$; Fig. 1c). Individuals who had been breastfed at some point as infants

[1]Department of Microbiology and Immunology, 1500 W. Medical Center, University of Michigan, Ann Arbor, Michigan 48109, USA.

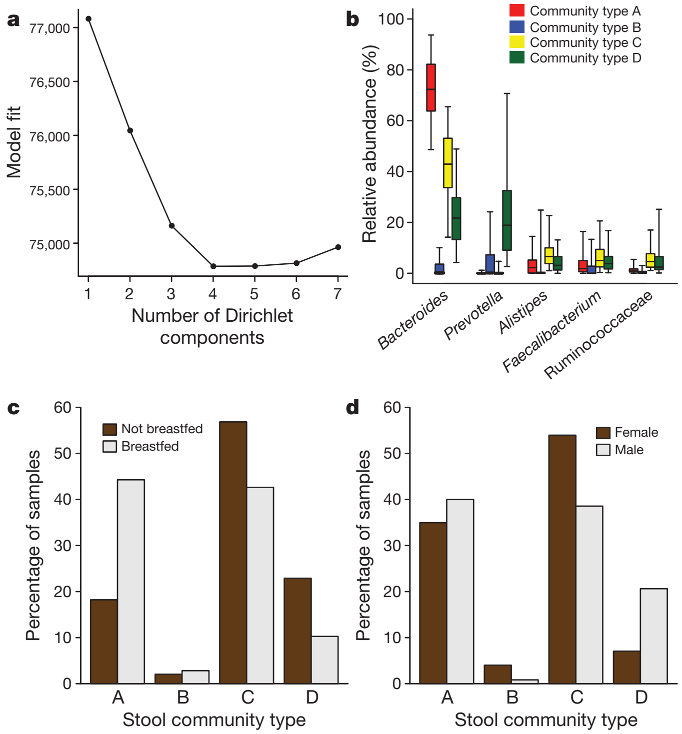


Figure 1 | Analysis of stool samples reveals four community types. **a**, Fitting the genera-level relative abundance data from 597 stool samples to Dirichlet multinomial mixture models provided support for four types when using the Laplace approximation to the negative log model evidence. **b**, The relative abundance of the most abundant genera in the samples assigned to each of the types (the boxes represent the interquartile range and the error bars represent the 95% confidence interval; n (community type A) = 221; n (community type B) = 15; n (community type C) = 80; n (community type D) = 281). **c**, **d**, There were significant associations between stool community types (n = 287 unique individuals) and whether the subject was breastfed as an infant (**c**; median $P = 1 \times 10^{-4}$) and their gender (**d**; median $P = 4 \times 10^{-4}$).

were 2.4-times more likely to belong to community type A, and those who were not breastfed were 2.2-times more likely to belong to community type D. Gender was associated with community types identified in the stool ($P = 4 \times 10^{-4}$; Fig. 1d), tongue ($P = 2 \times 10^{-3}$; Extended Data Fig. 2a), right retroauricular crease ($P = 9 \times 10^{-5}$; Extended Data Fig. 2b), and right antecubital fossa ($P = 3 \times 10^{-5}$; Extended Data Fig. 2c). For example, men were 3.0-times more likely than women to harbour stool community type D (Fig. 1b). Whether a woman had a baccalaureate degree had a strong association with the community types observed within the vaginal introitus ($P = 2 \times 10^{-3}$; Extended Data Fig. 3a), mid vagina ($P = 8 \times 10^{-4}$; Extended Data Fig. 3B), and posterior fornix ($P = 4 \times 10^{-4}$; Extended Data Fig. 3C). At each of these sites, women with a baccalaureate degree were more likely to be dominated by *Lactobacillus* (type E) and those without a baccalaureate degree were likely to have very low levels of *Lactobacillus* and moderate abundances of *Atopobium*, *Prevotella*, *Bifidobacterium* and unclassified members of the Firmicutes (type D). Together, our analysis indicates that an individual's life-history characteristics can be associated with their microbiome composition.

The second important observation that we identified was that the community type at one body site was predictive of the community type at another body site. Previously, cross-body site comparisons were made by calculating the ecological distance between samples collected at different body sites based on the taxonomic composition of those communities[3]. Our approach allowed us to identify similar associations within a body region (for example, oral, skin, vagina), but also allowed us to detect associations between communities that had very different taxonomic compositions. Community type membership was correlated among sites within the oral cavity, in the vagina, and between the left and right antecubital fossa and the left and right retroauricular crease (Fig. 2). Surprisingly, stool samples showed a significant association with samples from within the oral cavity; the strongest association was with the community types observed in saliva ($P = 10^{-3}$; Extended Data Table 3). Saliva was dominated by members of the *Prevotella*, *Streptococcus*, Pasteurellaceae, *Veillonella* and *Fusobacterium*; among these taxa, only *Prevotella* were abundant in the stool communities. Individuals with stool community type D, which had the highest level of *Prevotella*, were 2.1-times more likely to harbour saliva community types A and C, which were also high in *Prevotella* relative to saliva community types B and D. Stool community types A and C, which had low levels of *Prevotella*, were less likely to co-occur with saliva community types A and C (Extended Data Table 3). These results are intriguing because they suggest that although the oral and stool communities share little taxonomic resemblance, oral bacterial populations seed the gut, and those populations experience the ecological environment of the gut to give rise to consistent community types by the time they reach the stool.

Aside from life-history characteristics and inoculation from other body sites, the structure of the human microbiome is probably shaped by an individual's recent interactions with their environment, diet, medications, and overall health. We quantified the stability of each community type at every body site by estimating the probability that the type would change between sampling visits (Fig. 3a). The most stable body sites were in the stool and vagina and the least stable site was the supragingival plaque. Among the four stool community types, type D was the most stable followed by types A, C and B (Fig. 3b). Unfortunately, the metadata describing changes in health or lifestyle are unable to provide us with an explanation for why community types change.

The human microbiome is a complex ecosystem that varies considerably across the body and between individuals. This study demonstrates that given the myriad permutations of genetics, life histories, behaviours, environments and exposures, an individual's microbiome is an emergent property whereby a potentially limitless number of microbial community structures can be distilled into a finite number of types. Knowledge of

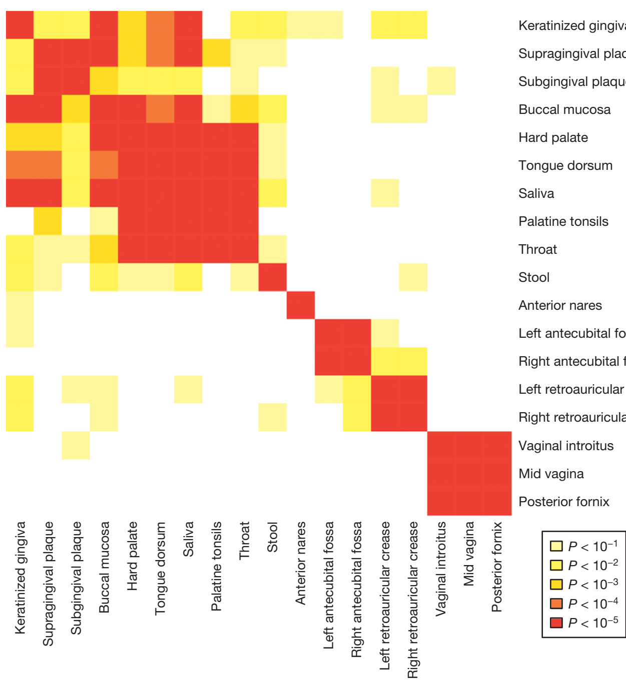


Figure 2 | Community-type associations are strongest within a body region, but also exist between stool and the oral cavity. Heat-map colours represent the magnitude of the median P value for the comparison of community type membership using Fisher's exact test. Median P values are found in the Source Data.

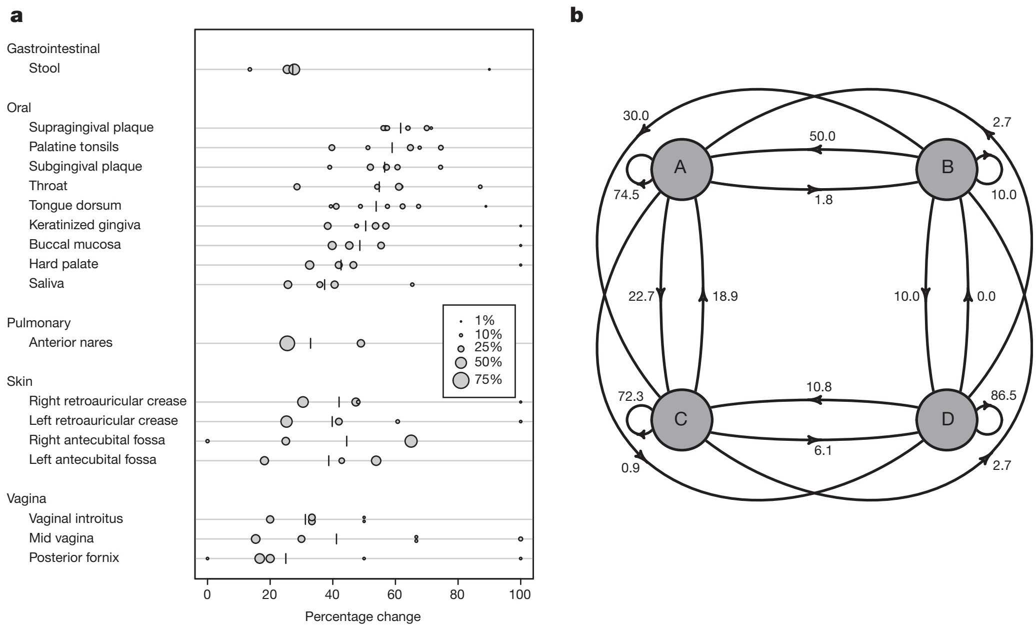


Figure 3 | Dynamics of community types at various body sites indicates that community type stability is correlated with the diversity of the community type. **a**, The community types at each body site differ in the fraction of samples that change their community type membership between visits. (Size of circles represents percentage of samples that affiliated with each community type and the vertical line represents the weighted average.) **b**, Rate of change between stool community types (*n* (community type A) = 221; *n* (community type B) = 15; *n* (community type C) = 80; *n* (community type D) = 281). The numbers on directed edges indicate the percentage of samples that changed community types.

the factors that affect one's community type profile will be critical as they continue to be associated with predisposition to diseases. Furthermore, understanding why community types change will be useful in developing therapies that can alter one's community type using pre- and probiotics, faecal transplants, or antibiotics. Given the varying levels of flux between community types at different body sites, it is remarkable that we were still able to detect life-long legacy effects on the microbiome, such as whether the subject was ever breastfed as an infant. This result could represent a true long-term impact of breastfeeding on the microbiome or it could represent the effect of the individual's childhood environment or care. The result raises the possibility that there may be other legacy effects on the microbiome, such as duration of breastfeeding, mode of birth, level of early antibiotic exposure, and childhood disease[18–20]. The four gender-based associations are intriguing and support previous studies showing that men and women have different skin communities[21] and that autoimmune diseases may be mediated via the microbiome and hormonal differences[22]. The association between one's level of education and their vaginal microbiome type is less clear; it is most likely that a baccalaureate degree represents a composite variable of numerous factors known to affect the vaginal microbiome, including race/ethnicity, sexual behaviour and socioeconomic class. Regardless, that such considerable variation was observed among a population of healthy women supports the observation that there is no single normal vaginal microbiome[23]; this is probably true for every body site. Looking forward, prospective studies that include individuals with varied levels of health and varied backgrounds (study groups that are more representative of society) are needed to achieve a better understanding of the mechanisms of change in community types as well as to provide more details about correlations between community type and life-history factors such as genetics, age, diet, health status, and environment (that is, rural or urban). Furthermore, future prospective studies with a longitudinal component need to control for the time between samplings and perhaps synchronize sampling with host physiology (for example, menses). Perhaps most exciting is the prospect that community types may be associated with complex diseases such as bacterial vaginosis, periodontitis, cancer, and diabetes where it has not been possible to establish a causative relationship between a specific bacterium and the disease.

METHODS SUMMARY

The Human Microbiome Project carried out three phases of sequencing the 16S rRNA gene and we obtained the unprocessed data for the V35 region from the NCBI Short Read Archive (SRA): the Clinical Pilot Project (accession SRP002012), Phase I (accession SRP002395) and Phase II (accession SRP002860). The Clinical Pilot Project and Phase I data sets have been described previously[2,3]. The metadata and clinical data associated with the samples from the subjects were obtained from dbGap (accession phs000228.v3.p1). The 16S rRNA gene sequence curation pipeline was implemented using the mothur software package (http://nbviewer.ipython.org/gist/pschloss/9815766/notebook.ipynb)[24,25].

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 2 September 2013; accepted 20 February 2014.
Published online 16 April 2014.

1. Peterson, J. *et al.* The NIH Human Microbiome Project. *Genome Res.* **19,** 2317–2323 (2009).
2. The Human Microboime Consortium. A framework for human microbiome research. *Nature* **486,** 215–221 (2012).
3. The Human Microboime Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* **486,** 207–214 (2012).
4. Holmes, I., Harris, K. & Quince, C. Dirichlet multinomial mixtures: generative models for microbial metagenomics. *PLoS ONE* **7,** e30126 (2012).
5. Aagaard, K. *et al.* The Human Microbiome Project strategy for comprehensive sampling of the human microbiome and why it matters. *FASEB J.* **27,** 1012–1022 (2013).
6. Turnbaugh, P. J. *et al.* A core gut microbiome in obese and lean twins. *Nature* **457,** 480–484 (2009).
7. Costello, E. K. *et al.* Bacterial community variation in human body habitats across space and time. *Science* **326,** 1694–1697 (2009).
8. Koren, O. *et al.* A guide to enterotypes across the human body: meta-analysis of microbial community structures in human microbiome datasets. *PLOS Comput. Biol.* **9,** e1002863 (2013).
9. Arumugam, M. *et al.* Enterotypes of the human gut microbiome. *Nature* **473,** 174–180 (2011).
10. Wu, G. D. *et al.* Linking long-term dietary patterns with gut microbial enterotypes. *Science* **334,** 105–108 (2011).
11. Quince, C. *et al.* The impact of Crohn's disease genes on healthy human gut microbiota: a pilot study. *Gut* **62,** 952–954 (2013).
12. Brotman, R. M. *et al.* Association between *Trichomonas vaginalis* and vaginal bacterial community composition among reproductive-age women. *Sex. Transm. Dis.* **39,** 807–812 (2012).
13. Gajer, P. *et al.* Temporal dynamics of the human vaginal microbiota. *Sci. Transl. Med.* **4,** 132ra152 (2012).

14. Ravel, J. *et al.* Vaginal microbiome of reproductive-age women. *Proc. Natl Acad. Sci. USA* **108** (suppl. 1), 4680–4687 (2011).
15. Statnikov, A. *et al.* Microbiomic signatures of psoriasis: feasibility and methodology comparison. *Sci. Rep.* **3,** 2620 (2013).
16. Arumugam, M. *et al.* Addendum: Enterotypes of the human gut microbiome. *Nature* **506,** 516 (2014).
17. Moeller, A. H. *et al.* Chimpanzees and humans harbour compositionally similar gut enterotypes. *Nature Commun.* **3,** 1179 (2012).
18. Palmer, C., Bik, E. M., Digiulio, D. B., Relman, D. A. & Brown, P. O. Development of the human infant intestinal microbiota. *PLoS Biol.* **5,** e177 (2007).
19. Koenig, J. E. *et al.* Succession of microbial consortia in the developing infant gut microbiome. *Proc. Natl Acad. Sci. USA* **108** (suppl. 1), 4578–4585 (2011).
20. Pantoja-Feliciano, I. G. *et al.* Biphasic assembly of the murine intestinal microbiota during early development. *ISME J.* **7,** 1112–1115 (2013).
21. Fierer, N., Hamady, M., Lauber, C. L. & Knight, R. The influence of sex, handedness, and washing on the diversity of hand surface bacteria. *Proc. Natl Acad. Sci. USA* **105,** 17994–17999 (2008).
22. Markle, J. G. *et al.* Sex differences in the gut microbiome drive hormone-dependent regulation of autoimmunity. *Science* **339,** 1084–1088 (2013).
23. Ma, B., Forney, L. J. & Ravel, J. Vaginal microbiome: rethinking health and disease. *Annu. Rev. Microbiol.* **66,** 371–389 (2012).
24. Schloss, P. D. *et al.* Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75,** 7537–7541 (2009).
25. Schloss, P. D., Gevers, D. & Westcott, S. L. Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PLoS ONE* **6,** e27310 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank J. Crabtree of the HMP Data Analysis and Coordination Center for his assistance in obtaining the sequencing and metadata files. The analysis described in this study was supported by grants from the National Institutes of Health (R01HG005975, R01GM099514 and P30DK034933).

Author Contributions T.D. and P.D.S. designed and executed the analysis and prepared the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.D.S. (pschloss@umich.edu).

METHODS

Sequence analysis pipeline. The Human Microbiome Project carried out three phases of sequencing the 16S rRNA gene, which were performed using the 454 Titanium sequencing platform. We obtained the unprocessed sff files for the V35 region from the NCBI Short Read Archive (SRA) for each of these phases: the Clinical Pilot Project (accession SRP002012), Phase I (accession SRP002395) and Phase II (accession SRP002860). The Clinical Pilot Project and Phase I data sets have been described previously[2,3]. The sequencing was performed by sequencing from the 3′ to the 5′ end of the 16S rRNA gene[26]. Although the V13 and V69 regions were also sequenced by the HMP sequencing centres, the number of data sets generated for those regions was considerably smaller than was obtained for the V35 region. The 16S rRNA gene sequence curation pipeline was implemented using the mothur software package[24,25]. This approach has been shown to result in a sequencing error rate of 0.02%[25]. Briefly, flowgrams were extracted from the sff files and any that had more than one mismatch to the barcode, more than two mismatches to the primer, had fewer than 450 flows, contained homopolymers longer than 8 nucleotides, or contained an ambiguous base call were culled. The flows for each sequencing run were trimmed to 450 flows and de-noised separately using the PyroNoise algorithm as implemented within mothur[27]. The de-noised sequences were then aligned against a customized reference alignment based on the SILVA database using the NAST algorithm implemented within mothur[28]. The customized database included small subunit rRNA sequences from bacteria, archaea, eukarya, chloroplasts and mitochondria. Sequences that did not align to the predicted V35 region were culled from further analysis and the alignments were trimmed so that the sequences fully overlapped the same alignment coordinates[29,30]. These sequences were then subjected to a pre-clustering step that first sorted the sequences by their abundance within each sample and then clustered sequence abundances together if a sequence was within 2 nucleotides of a more abundant sequence[25]. Treating each sample separately, we interrogated each sequence for the presence of chimaeras using the *de novo* UChime chimaera detection algorithm[31]. Once chimaeric sequences were culled from the data sets, the sequences were classified using the naive Bayesian Classifier trained against a customized version of the RDP training set (version 9) as implemented within mothur[32]. The training set was customized by supplementing sequences derived from chloroplasts, mitochondria and members of the Eukarya. The reference sequences were trimmed to only include the V35 region of the 16S rRNA gene. We required a minimum classification confidence score of 80% and used 1,000 pseudo-bootstrap iterations. Because the PCR target was bacterial 16S rRNA gene sequences, we culled any sequences that classified as being derived from archaea, eukarya, mitochondria, chloroplasts or sequences that could not be classified to a kingdom with at least 80% confidence. The taxonomy of the remaining sequences was used to assign the sequences to genus-level phylotypes. Those sequences without a genus-level classification were assigned to a phylotype represented by the lowest level taxonomy with a confidence score of at least 80%. This allowed us to create a table of counts for the number of times each genus-level phylotype was observed in each sample. As some samples were sequenced multiple times to obtain additional sequence data, we pooled replicate sequencing runs to create a single sample. Samples with fewer than 1,000 reads were removed from further analysis and all samples were either sub-sampled or rarefied ($n = 1,000$ iterations) to 1,000 reads to perform subsequent analyses. Sub-sampling and rarefaction were necessary to limit the effects of differential sampling that are known to affect alpha and beta diversity metrics and differentially increase the representation of PCR and sequencing artefacts in data sets.

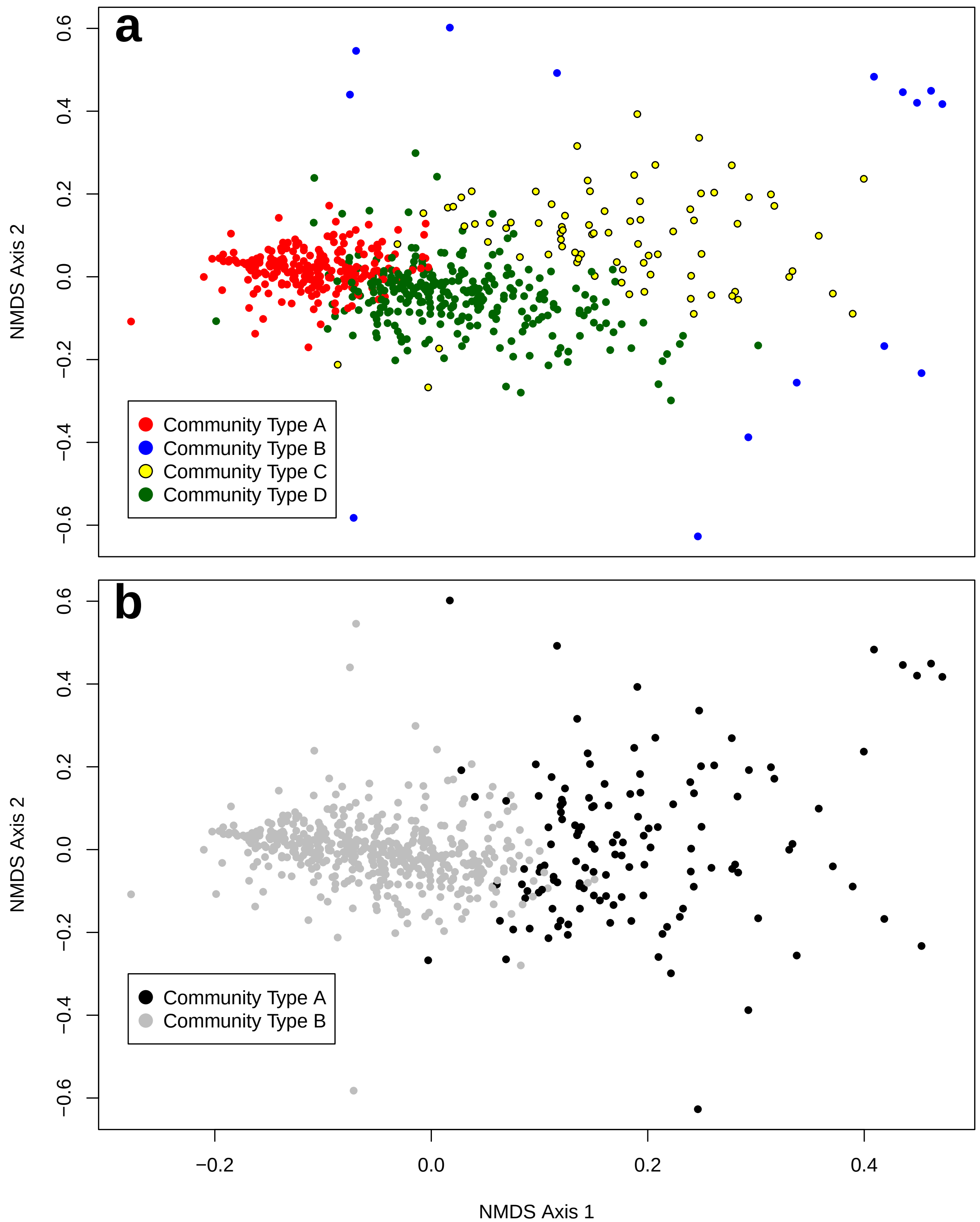
Assignment to community types. The table of counts was partitioned according to the 18 body sites. Because the communities were similar and we wanted to use the maximum number of samples per body site when assigning samples to community types, we pooled the three vaginal body sites (that is, vaginal introitus, mid-vagina, and posterior fornix) into one vaginal data set, the two antecubital fossa sites (that is, left and right) into one antecubital fossa data set, and the two retroauricular crease sites (that is, left and right) into one retroauricular crease data set. The resulting 14 tables were used as input to partition the samples according to community types at each body site using the Dirichlet multinomial mixture model[4]. We selected the number of community types at each body site by selecting the number of components that gave the minimum Laplace approximation to the negative log model evidence. Samples were assigned to their community type based on the maximum posterior probability. For all body sites, between 89.2% and 99.7% of the samples had a posterior probability of at least 0.90. The mean abundance and 95% confidence interval predicted by the model are provided for each body site.

Selection of metadata. A large amount of metadata and clinical data were collected for each of the samples and subjects[5]. We obtained the most recent version of these data from dbGap (accession phs000228.v3.p1). Because of the uniformity and healthy nature of the cohort, a number of the clinical data fields could not be included in our analysis. Furthermore, there was evidence that several variables were collected from subjects in one city but not the other (see Supplementary Information for a discussion of the difficulties in analysing the city of origin data). We interrogated, a priori, the categorical metadata to identify those variables where we were able to identify at least 10 instances of the condition and that was represented in the subjects from both cities. In addition, to increase the number of variables under consideration, we pooled responses. For example, there were 13 categories for country of birth with only one of those having more than 10 respondents (US/Canada; $n = 260$). In this case we pooled the other responses to create a non-US/Canada group ($n = 40$). We used a similar pooling strategy for parents' country of origin, meat eaters/vegetarians, number of children the subject had given birth to, occupation, and level of education. The data available through dbGap partitioned medications into broad categories and indicated whether the subject was using the medication at the time of the visit. This created three classes of subjects. The first never used the medication during the study, the second class used the medication for one or two of the visits, but not all of their samples, and the third class used the medication for all of their visits. For the purpose of correlating medication usage with community type, we used the data for the first and third classes of subjects and ignored the second. The number of subjects in the second class was below 10 for each type of medication. For example, there were only 8 subjects that had more than one visit and used antibiotics within 30 days of any of their visits. Because of the general paucity of subjects in this category of medication users, it was not possible to associate medication usage with changes in community type. Finally, we converted the subjects' body mass index (BMI) into categories of normal (18.5–25), overweight (25–30) and obese (>30); there were no underweight subjects (<18.5). The resulting list of categorical clinical data that were considered is provided in Extended Data Table 1. In addition to these categorical data, we also had access to continuous clinical metadata for each of the subjects. These included their age, BMI, pulse and blood pressure. A summary of these data is provided in Extended Data Table 1.

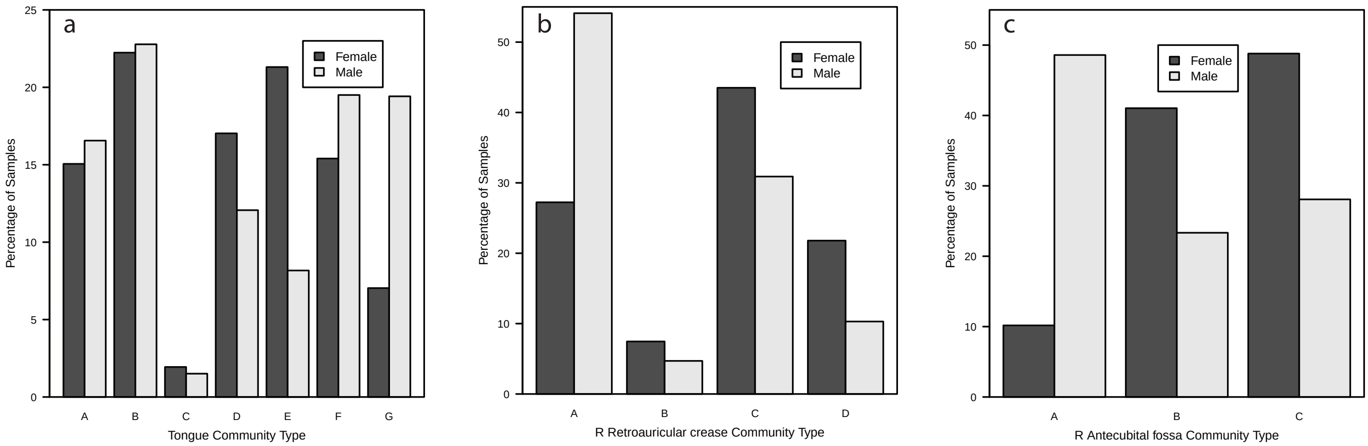
Tests of association. Because individuals provided up to three samples it would have been arbitrary to select one visit from each subject (for example, the first visit) on which to base our analyses. For the categorical metadata associations, we performed an iterative procedure where we selected a single visit for each individual's body site and tested the association between the community type at the body site and the metadata using Fisher's exact test. For the continuous metadata we performed a similar procedure except we tested the association between the community type at the body site and the metadata using analysis of variance. Finally, to test associations across the body we performed a randomization procedure where each iteration consisted of selecting one visit for each individual and then testing for inter-body site associations using the Fisher exact test. We performed 1,000 iterations and we calculated the percentage of iterations that each variable was significant according to the Benjamini–Hochberg step-up procedure that we used to limit the false discovery rate to 5%. We report the median *P* value and the percentage of iterations that were significant.

A complete description of the analysis pipeline including scripts, mothur commands, and intermediate files are available at http://nbviewer.ipython.org/gist/pschloss/9815766/notebook.ipynb.

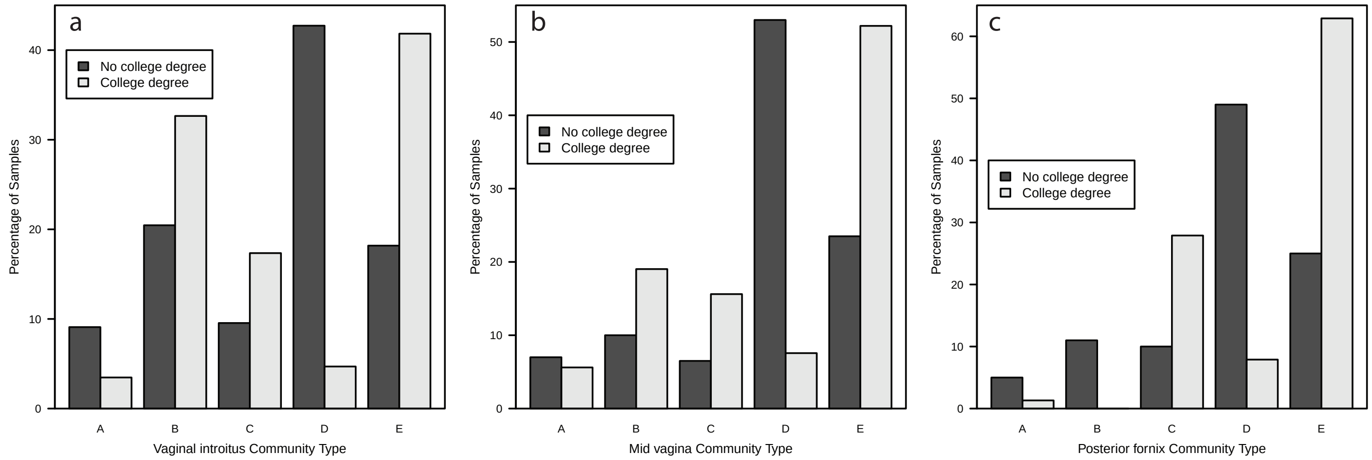
26. Jumpstart Consortium Human Microbiome Project Data Generation Working Group. Evaluation of 16S rDNA-based community profiling for human microbiome research. *PLoS ONE* **7,** e39315 (2012).
27. Quince, C., Lanzen, A., Davenport, R. J. & Turnbaugh, P. J. Removing noise from pyrosequenced amplicons. *BMC Bioinform.* **12,** 38 (2011).
28. Schloss, P. D. A high-throughput DNA sequence aligner for microbial ecology studies. *PLoS ONE* **4,** e8230 (2009).
29. Schloss, P. D. Secondary structure improves OTU assignments of 16S rRNA gene sequences. *ISME J.* **7,** 457–460 (2013).
30. Pruesse, E. *et al.* SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* **35,** 7188–7196 (2007).
31. Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C. & Knight, R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27,** 2194–2200 (2011).
32. Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73,** 5261–5267 (2007).



Extended Data Figure 1 | Comparison of community type assignments for non-metric dimensional scaling (NMDS) ordination of Jensen–Shannon divergence values between stool samples using DMM (a) and PAM-based clustering (b). The stress computed for this ordination was 0.19 and the R^2 between the input distance matrix and the distance matrix calculated between the points in the ordination was 0.90.



Extended Data Figure 2 | The frequency of community types for body sites where there was a significant association with the subject's gender. **a**–**c**, Percentage of female and male tongue communities that affiliated with each of the tongue (**a**; $n = 288$ unique individuals; median $P = 2 \times 10^{-3}$), right retroauricular crease (**b**; $n = 268$ unique individuals; median $P = 9 \times 10^{-5}$) and right antecubital fossa community types (**c**; $n = 136$ unique individuals; median $P = 3 \times 10^{-5}$).



Extended Data Figure 3 | The frequency of vaginal community types among women with and without a college degree. a–c, Percentage of women with and without a college degree whose vaginal communities affiliated with the vaginal introitus (**a**; $n = 74$ unique individuals; median $P = 2 \times 10^{-3}$), mid-vagina (**b**; $n = 64$ unique individuals; median $P = 8 \times 10^{-4}$) and posterior fornix (**c**; $n = 61$ unique individuals; median $P = 4 \times 10^{-4}$) community types.

Extended Data Table 1 | Most common characteristics of the individuals included in the HMP healthy cohort

Categorical data	Number of Individuals (Total=300)
Sampled in Houston / St. Louis	150 / 150
Female / Male	151 / 149
Born in US or Canada	260
Mother born in US or Canada	227
Father born in US or Canada	229
Hispanic, Latino, or Spanish	32
Asian	34
Black	21
White	243
Ever breastfed as infant	198 (Forgot=33, NA=23)
Eats meat at least once a week	261 (NA=23)
Occupation: Student	150
Had given birth at least once	26 (NA=156)
College educated	210 (NA=16)
Had dental insurance	265 (NA=16)
Had health insurance	217 (NA=16)
Tobacco user	19
Chronic use of antidepressants	22 (Tr=9)
Chronic use of antihistamines	9 (Tr=13)
Chronic use of hormonal contraceptives	72 (Tr=15; NA=149)
Chronic use of vitamins or supplements	34 (Tr=10)
Normal BMI	178
Overweight BMI	88
Obese BMI	34

Continuous data	Median (Min-Max)
Age	25 (18-40)
BMI	24 (19-34)
Pulse	71 (42-100)
Diastolic pressure	71 (50-98)
Systolic pressure	119 (91-151)
pH – Vaginal introitus	4.4 (3.3-6.5)
pH – Posterior fornix	4.0 (3.2-7.0)

NA, data was not collected; Forgot, the subject could not recall; Tr, the medication was used transiently throughout the two or three sampling events.

Extended Data Table 2 | Comparison of PAM- and DMM-based approaches to assigning samples to community types

	PAM-based using SI Index			PAM-based using CH Index			DMM-based	
Body site	Clusters	SI Index	Laplace	Clusters	CH Index	Laplace	Clusters	Laplace
Antecubital fossa	2	0.34	84858.4	2	114.3	84858.4	3	83302.1
Anterior nares	3	0.32	52136.3	2	153.5	51864.3	2	51532.0
Buccal mucosa	2	0.23	65643.3	4	166.2	64968.1	4	64588.8
Hard palate	2	0.28	72686.9	4	208.4	71573.8	4	71436.9
Keratinized gingiva	2	0.38	51392.2	2	323.8	51392.2	5	50605.3
Palatine tonsils	2	0.27	82655.3	2	237.7	82655.3	6	81446.7
Retroauricular crease	3	0.51	95797.5	3	719.9	95797.5	5	94673.5
Saliva	2	0.19	81261.3	2	120.4	81261.3	4	80656.1
Stool	2	0.40	76228.5	2	194.0	76228.5	4	74785.6
Subgingival plaque	2	0.25	90876.7	2	249.4	90876.7	5	89672.2
Supragingival plaque	2	0.24	78982.0	2	217.6	78982.0	6	78357.1
Throat	2	0.26	79238.0	2	177.8	79238.0	5	78052.8
Tongue dorsum	2	0.33	71442.3	2	293.2	71442.3	7	69923.0
Vagina	10	0.57	32407.3	2	205.7	32150.8	5	31209.5

Extended Data Table 3 | Average contingency table of stool and saliva community types

	Saliva A	Saliva B	Saliva C	Saliva D
Stool A	0.101	0.140	0.104	0.044
Stool B	0.003	0.000	0.002	0.021
Stool C	0.136	0.173	0.107	0.027
Stool D	0.048	0.024	0.052	0.017

The median P value from a Fisher's exact test was 1×10^{-3}.

doi:10.1038/nature13160

T-cell activation by transitory neo-antigens derived from distinct microbial pathways

Alexandra J. Corbett[1]*, Sidonia B. G. Eckle[1]*, Richard W. Birkinshaw[2]*, Ligong Liu[3]*, Onisha Patel[2], Jennifer Mahony[4], Zhenjun Chen[1], Rangsima Reantragoon[1], Bronwyn Meehan[1], Hanwei Cao[1], Nicholas A. Williamson[5], Richard A. Strugnell[1], Douwe Van Sinderen[4,6], Jeffrey Y. W. Mak[3], David P. Fairlie[3,7]*, Lars Kjer-Nielsen[1]*, Jamie Rossjohn[2,8,9]* & James McCluskey[1]*

T cells discriminate between foreign and host molecules by recognizing distinct microbial molecules, predominantly peptides and lipids[1–4]. Riboflavin precursors found in many bacteria and yeast also selectively activate mucosal-associated invariant T (MAIT) cells[5,6], an abundant population of innate-like T cells in humans[7–9]. However, the genesis of these small organic molecules and their mode of presentation to MAIT cells by the major histocompatibility complex (MHC)-related protein MR1 (ref. 8) are not well understood. Here we show that MAIT-cell activation requires key genes encoding enzymes that form 5-amino-6-D-ribitylaminouracil (5-A-RU), an early intermediate in bacterial riboflavin synthesis. Although 5-A-RU does not bind MR1 or activate MAIT cells directly, it does form potent MAIT-activating antigens via non-enzymatic reactions with small molecules, such as glyoxal and methylglyoxal, which are derived from other metabolic pathways. The MAIT antigens formed by the reactions between 5-A-RU and glyoxal/methylglyoxal were simple adducts, 5-(2-oxoethylideneamino)-6-D-ribitylaminouracil (5-OE-RU) and 5-(2-oxopropylideneamino)-6-D-ribitylaminouracil (5-OP-RU), respectively, which bound to MR1 as shown by crystal structures of MAIT TCR ternary complexes. Although 5-OP-RU and 5-OE-RU are unstable intermediates, they became trapped by MR1 as reversible covalent Schiff base complexes. Mass spectra supported the capture by MR1 of 5-OP-RU and 5-OE-RU from bacterial cultures that activate MAIT cells, but not from non-activating bacteria, indicating that these MAIT antigens are present in a range of microbes. Thus, MR1 is able to capture, stabilize and present chemically unstable pyrimidine intermediates, which otherwise convert to lumazines, as potent antigens to MAIT cells. These pyrimidine adducts are microbial signatures for MAIT-cell immunosurveillance.

MAIT-cell antigens were previously identified from *Salmonella typhimurium* (strain SL1344) supernatant[5]. Negative-mode electrospray ionization–time-of-flight mass spectrometry (ESI–TOF-MS) analysis of MR1-bound ligands from *S. typhimurium* revealed a ligand with a mass to charge (*m*/*z*) ratio of 329.11, matching a potent MAIT-activating ligand identified during the chemical synthesis of reduced 6-hydroxymethyl-8-D-ribityllumazine (rRL-6-CH_2OH) (ref. 5). Although this ligand was identified biochemically, its origin was puzzling, as it is not described in the riboflavin synthesis pathway. We therefore took a genetic approach to evaluate whether the riboflavin pathway supplied the MAIT-cell ligands.

We examined the capacity of bacterial mutants of the riboflavin pathway to activate MAIT cells (Fig. 1a). In some bacterial species, including *Lactococcus lactis*, the genes necessary for riboflavin synthesis are grouped together in a single four-gene operon (RibGBAH), and are regulated by transcriptional repression of a 'riboswitch' via flavin mononucleotide and riboflavin[10]. Using *L. lactis*, we tested the ability of bacterial culture supernatant to activate Jurkat cells transduced with a MAIT T-cell antigen receptor (TCR) (Jurkat.MAIT) (Fig. 1b). Supernatant from wild-type *L. lactis* strain NZ9000 incubated with antigen-presenting cells expressing MR1 caused CD69 upregulation in Jurkat.MAIT cells (Fig. 1b). Addition of riboflavin during culture of NZ9000 abrogated Jurkat.MAIT-cell activation, consistent with negative regulation of the riboswitch and impaired production of the activating MAIT ligand (Fig. 1b). Next, three mutant strains of *L. lactis* were used: two riboflavin overproducers, CB013 and CB021, which produce riboflavin even in the presence of high riboflavin concentrations; and a RibA$^-$ strain, which contains a deletion in *ribA*, early in the riboflavin pathway[11]. The riboflavin overproducers activated Jurkat.MAIT cells when grown with or without exogenous riboflavin, whereas there was no MAIT-cell activation by supernatant from the RibA$^-$ strain (Fig. 1b). Accordingly, the riboflavin pathway is necessary and sufficient to produce natural MAIT-cell antigens.

Next we generated individual mutations in the four genes of the riboflavin operon in *L. lactis*. These were produced using the constitutive riboflavin overproducer strain CB013. Culture supernatants from bacteria with mutant riboflavin pathways were tested for activation of Jurkat.MAIT cells (Fig. 1c). The parental CB013 supernatant activated Jurkat.MAIT cells, whereas bacteria containing mutations in *ribA* or *ribG* did not activate the reporter cells under similar conditions (Fig. 1c). Neither *ribB* nor *ribH* mutations, which affect the pathway downstream of 5-A-RU, had any impact on Jurkat.MAIT activation (Fig. 1c). Moreover, whereas the *m*/*z* 329.11 species was undetectable in MR1 refolded with supernatant from the RibA and RibG mutants, it was captured by MR1 from the supernatants of the RibB and RibH mutants of *L. lactis* (Extended Data Fig. 1a). Moreover, culture supernatants from *S. typhimurium* SL1344 with mutated RibD and RibH (RibD+H) did not furnish a detectable *m*/*z* 329.11 species that bound MR1, and could not activate Jurkat.MAIT cells (Extended Data Fig. 1b, c). However, antigen was detectable in the complemented, activating RibD+H SL1344 mutant (Extended Data Fig. 1b, c). MR1-restricted ligands were not detected from the supernatant of *Enterococcus faecalis*, which neither possesses the riboflavin pathway nor activates MAIT cells (Extended Data Fig. 1d and data not shown). Analysis of MR1-bound ligands from another MAIT-activating strain, *Escherichia coli* (DH5α strain) also revealed a ligand with an *m*/*z* ratio of 329.11 (Extended Data Fig. 1e). These data are consistent with MAIT-activating ligands, from a number of bacterial sources, being derived through an unknown mechanism from 5-A-RU.

A key precursor step in riboflavin biosynthesis is the condensation of 5-A-RU (**1**) with 3,4-dihydroxy-2-butanone-4-phosphate (**2a**) to

[1]Department of Microbiology and Immunology, Peter Doherty Institute for Infection and Immunity, University of Melbourne, Parkville, Victoria 3010, Australia. [2]Department of Biochemistry and Molecular Biology, School of Biomedical Sciences, Monash University, Clayton, Victoria 3800, Australia. [3]Division of Chemistry and Structural Biology, Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland 4072, Australia. [4]School of Microbiology, University College Cork, Cork, Ireland. [5]The Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Parkville, Victoria 3010, Australia. [6]Alimentary Pharmabiotic Centre, University College Cork, Cork, Ireland. [7]Australian Research Council Centre of Excellence in Advanced Molecular Imaging, The University of Queensland, Brisbane, Queensland 4072, Australia. [8]Australian Research Council Centre of Excellence in Advanced Molecular Imaging, Monash University, Clayton, Victoria 3800, Australia. [9]Institute of Infection and Immunity, Cardiff University, School of Medicine, Heath Park, Cardiff CF14 4XN, UK.
*These authors contributed equally to this work.

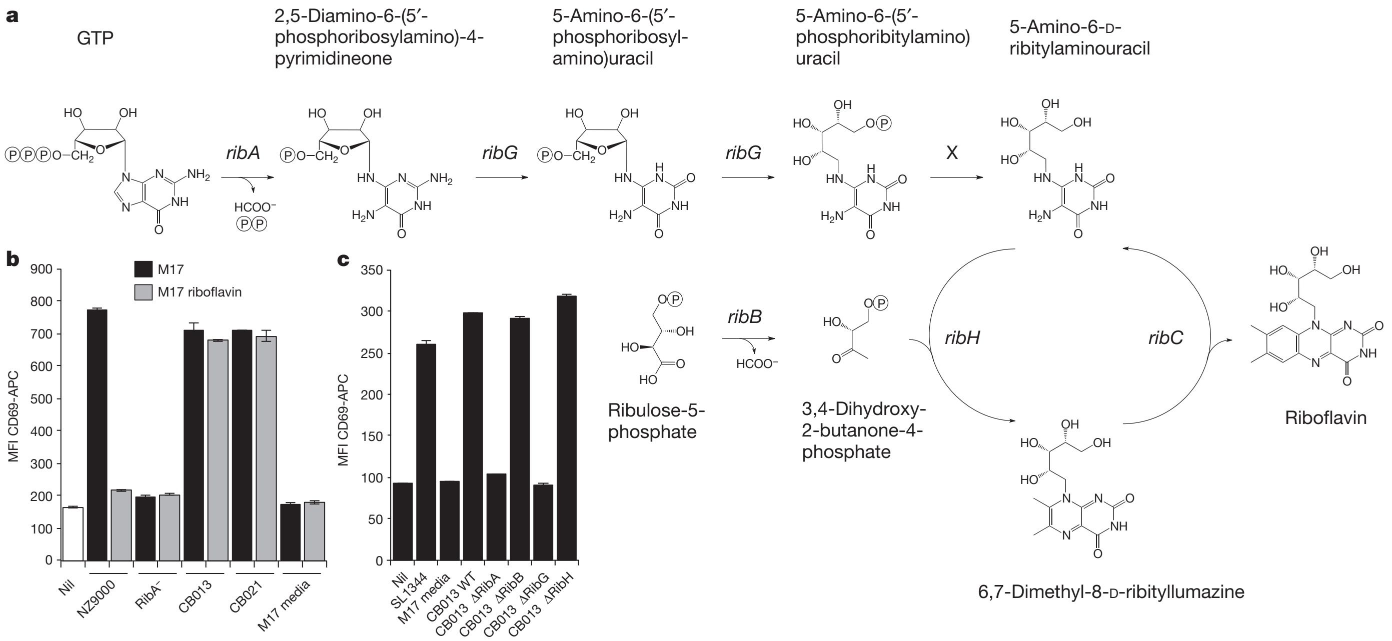


Figure 1 | The riboflavin pathway furnishes ligands that activate MAIT cells. **a**, Riboflavin biosynthesis pathway. *ribH*, lumazine synthase; X, hypothetical phosphatase. **b**, Cells were incubated overnight with filtered supernatant S/N from *L. lactis* NZ9000 (wild type), RibA$^-$ (RibA deletion mutation), CB013 and CB021 (riboflavin overproducers) overnight cultures plus or minus 3 μg ml^{-1} riboflavin, then stained with anti-CD3 monoclonal antibody coupled to PE and anti-CD69 monoclonal antibody coupled to APC. Mean fluorescence intensity (MFI) anti-CD69-APC for gated Jurkat.MAIT cells. Data are shown as mean ± standard error of the mean (s.e.m.). **c**, Cells were incubated overnight with 10 μl filtered, culture S/N from *Lactococcus* CB013 (deregulated riboflavin operon; wild type (WT)), CB013ΔRibA, CB013ΔRibB, CB013ΔRibG and CB013ΔRibH or *S. typhimurium*, then stained with anti-CD3-PE and anti-CD69-APC. MFI anti-CD69-APC for gated Jurkat.MAIT cells. Data are shown as mean ± s.e.m. Experiments were performed at least three times.

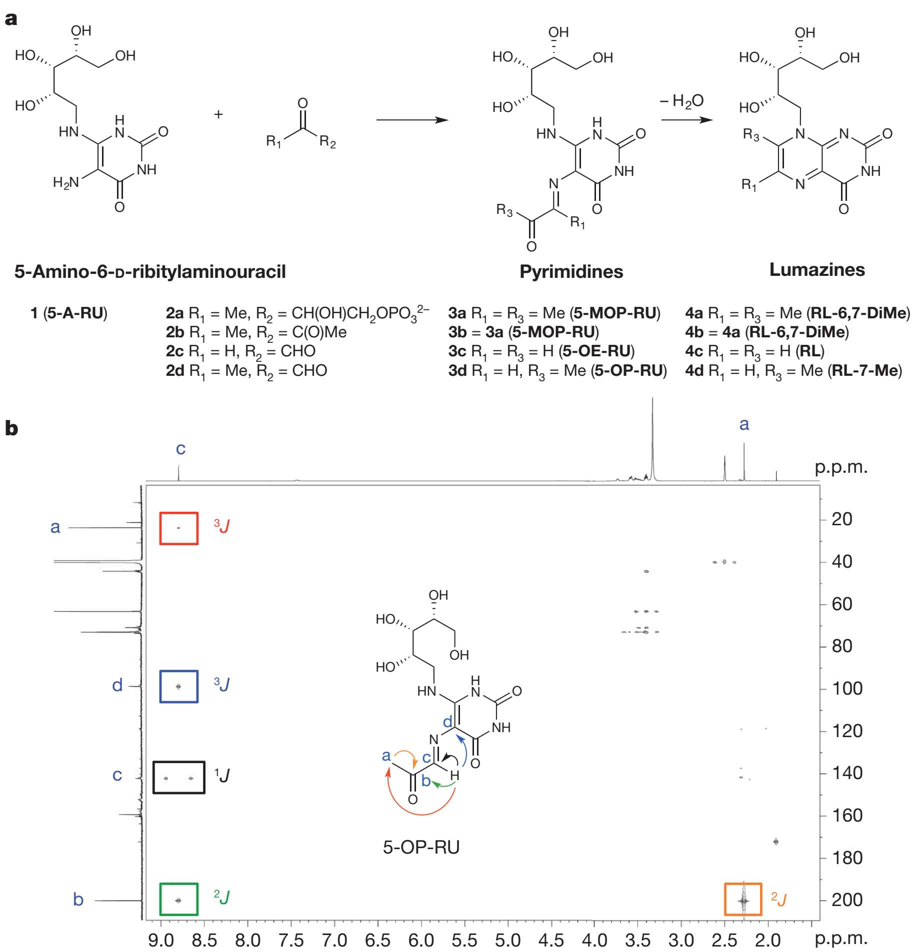


Figure 2 | Chemical formation of pyrimidines and lumazines from condensation of small metabolites with 5-A-RU. **a**, Series a, 5-A-RU (**1**) and 3,4-dihydroxy-2-butanone-4-phosphate (**2a**) form 5-MOP-RU (**3a**), and then RL-6,7-DiMe (**4a**). Series b–d, 5-A-RU (**1**) with butane-2,3-dione (**2b**), glyoxal (**2c**) and methylglyoxal (**2d**) forms 5-MOP-RU (**3b** = **3a**), 5-OE-RU (**3c**) and 5-OP-RU (**3d**), respectively, and then RL-6,7-DiMe (**4b** = **4a**), RL (**4c**) and RL-7-Me (**4d**), respectively. **b**, Two-dimensional NMR spectrum (heteronuclear multiple bond correlation (HMBC)) of isolated 5-OP-RU (**3d**) in DMSO-d_6 showing key ^{1}H–^{13}C long-range correlations that unambiguously characterize the imine adduct (**3d**), also identified in aqueous media (pH > 6). J refers to heteronuclear coupling through 1, 2 or 3 bonds.

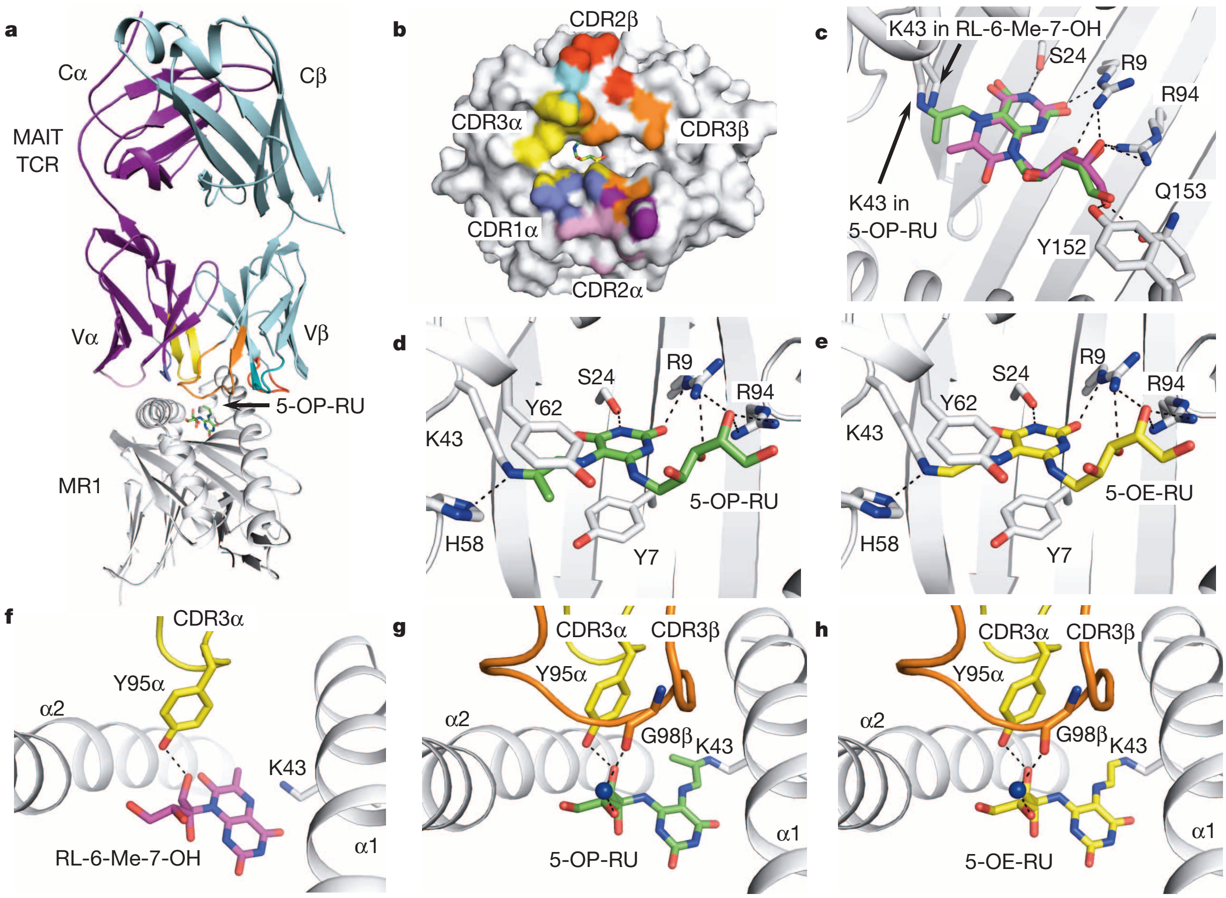


Figure 3 | Structural basis of MR1-binding and recognition of transitory MAIT-cell antigens. a–c, MAIT TCR–MR1–antigen docking (**a**), MAIT TCR footprint on MR1 surface (**b**) and 5-OP-RU and RL-6-Me-7-OH overlay (**c**). **d–h**, MR1 contacting 5-OP-RU (**d**) and 5-OE-RU (**e**); MAIT TCR contacting RL-6-Me-7-OH (**f**), 5-OP-RU (**g**) or 5-OE-RU (**h**). MR1 is depicted in grey, α-chain in purple, β-chain in cyan. 5-OP-RU is shown in green, 5-OE-RU in yellow, RL-6-Me-7-OH in magenta, CDR1α in slate, CDR2α in pink, CDR3α in yellow, CDR1β in teal, CDR2β in red and CDR3β in orange.

generate an intermediate 5-(1-methyl-2-oxopropylideneamino)-6-D-ribitylaminouracil (5-MOP-RU; **3a**), which readily undergoes ring closure with dehydration to form 6,7-dimethyl-8-D-ribityllumazine (RL-6,7-DiMe; **4a**)[12,13] (Fig. 2a), a biosynthesis that is catalysed by lumazine synthase (RibH). However, RL-6,7-DiMe can also be generated in the absence of lumazine synthase[13,14], suggesting that MAIT antigens might be formed through spontaneous reactions of 5-A-RU with other small molecules through non-enzymatic mechanisms (Fig. 2a). For example, butane-2,3-dione (**2b**), glyoxal (**2c**) and methylglyoxal (also known as pyruvaldehyde; **2d**) can represent by-products arising from a number of metabolic pathways, including glycolysis[15]. Their condensations with 5-A-RU (**1**) would respectively produce the pyrimidine adducts 5-MOP-RU (**3b** = **3a**), 5-OE-RU (**3c**) and 5-OP-RU (**3d**) *en route* to ribityllumazines RL-6,7-DiMe (**4b** = **4a**), 8-D-ribityllumazine (RL; **4c**) and 7-methyl-8-D-ribityllumazine (RL-7-Me; **4d**), respectively. We found that the initial adducts **3b–d** were formed almost immediately, but readily underwent dehydration upon ring closure to form the stable, isolatable compounds **4b–d** (Extended Data Figs 7 and 8), without the need for enzyme catalysis. Adducts **3b–d** (Fig. 2a) were especially unstable under acidic aqueous conditions (pH < 6), but we could detect them in solution under physiological conditions. We were able to synthesize **3d** in deuterated dimethylsulphoxide (DMSO-d_6), isolate it, unambiguously assign its solution structure by NMR spectroscopy (Fig. 2b and Extended Data Fig. 2), and examine its stability in aqueous media using liquid chromatography–mass spectrometry (LC–MS). At 37 °C and pH 6.8, adduct **3d** was clearly formed and had a half-life ($t_{1/2}$) of around 2 h at 65 μM. It was more stable at lower temperatures (for example, $t_{1/2}$ 14–15 h at 15 °C, pH 6.8–8.0, 65–250 μM) (Extended Data Fig. 3).

We considered whether butane-2,3-dione (**2b**), glyoxal (**2c**) and methylglyoxal (**2d**), in the presence of 5-A-RU (**1**), might react spontaneously to facilitate MR1 refolding. MR1-5-A-RU was undetectable when MR1 was refolded with 5-A-RU alone (data not shown), and the presence of 5-A-RU and butane-2,3-dione failed to yield any MR1–antigen complexes (data not shown). However, refolding of MR1 in the presence of 5-A-RU and either glyoxal or methylglyoxal led to a correctly folded MR1–antigen complex. To understand the basis for ligand selectivity by MR1, we determined the structures of the MAIT TCR in complex with MR1 and antigens formed from the condensation of 5-A-RU and either methylglyoxal or glyoxal[16] (Fig. 3a–h, Extended Data Table 1 and Extended Data Fig. 4). Surprisingly, the chemically unstable adducts 5-OP-RU (compound **3d**; Fig. 2a) and 5-OE-RU (compound **3c**, Fig. 2a) were observed bound to MR1 (Extended Data Fig. 4 and Fig. 3d, e). Both of these one-ring (pyrimidine) compounds were thus captured by MR1, despite being relatively unstable in the absence of MR1 and readily undergoing dehydrative cyclization to compounds **4d** and **4c**, respectively. The aromatic pyrimidine ring systems of 5-OP-RU and 5-OE-RU superposed on the corresponding ring from the bicyclic lumazine 7-hydroxy-6-methyl-8-D-ribityllumazine (RL-6-Me-7-OH) (ref. 16) (Fig. 3c). The creation of 5-OP-RU or 5-OE-RU generated an aliphatic moiety that burrowed into the MR1 cleft, within which the residual carbonyl group formed a Schiff base with Lys 43 of MR1 (Fig. 3d, e). This aliphatic moiety was also stabilized in the cleft by interactions with Tyr 7 and Tyr 62 (Fig. 3d, e). By contrast, RL-6-Me-7-OH was non-covalently bound within MR1 (Fig. 3f). Moreover, RL-6-Me-7-OH does not have the propensity to tautomerize into a single-ring pyrimidine system owing to its ability to form a very stable amide tautomer. Nevertheless, the ribityl

moieties of 5-OP-RU, 5-OE-RU and RL-6-Me-7-OH were all located in essentially identical positions within their respective complexes, each forming a hydrogen bond to Tyr 95α of the MAIT TCR[16] (Fig. 3f–h). Notably, 5-OP-RU and 5-OE-RU are relatively unstable in aqueous media and thus MR1 can capture and stabilize pyrimidine intermediates in the synthesis of lumazines.

Next we undertook ESI–TOF-MS to identify independently the chemical composition of the ligands captured within these refolded MR1–antigen complexes. For MR1 refolded with 5-A-RU and methylglyoxal, a single peak with a retention time of 8.9 min and *m*/*z* 329.11 matched a species that was captured by MR1 from *Salmonella* supernatant[5] and from the reaction mixture during synthesis of rRL-6-CH_2OH (ref. 5) (Extended Data Fig. 5a, b). This finding is consistent with the identification, within the crystal structure with MR1, of 5-OP-RU independently assembled from 5-A-RU and methylglyoxal (Fig. 3d), and supported by the NMR and kinetic characterization of 5-OP-RU in solution (Fig. 2b and Extended Data Figs 2, 3). Similarly, mass spectrometric analysis of MR1 refolded with the mixture of 5-A-RU and glyoxal revealed precursor and product *m*/*z* values (Extended Data Fig. 5a, b) consistent with identification of 5-OE-RU within the crystal structure of MR1 refolded with 5-A-RU and glyoxal (Fig. 3e, h). Furthermore, mass spectrometric analysis of MR1 refolded with 5-A-RU and ^{13}C-labelled glycolaldehyde yielded the expected *m*/*z* 317.10 precursor and 179.04 product ions, in agreement with the *m*/*z* 315.09 precursor and 177.04 product ions identified in MR1 refolded with 5-A-RU and glyoxal (Extended Data Fig. 5a, b).

We asked whether the activity with synthetic rRL-6-CH_2OH (ref. 5) might reflect capture by MR1 of a synthetic intermediate. The ligand captured by the mutant MR1(K43A) exposed to the reaction mixture generating rRL-6-CH_2OH (ref. 5) was identical, as identified by LC–MS and MS/MS analysis (*m*/*z* 329.11), to the MR1-bound antigen from either the *Salmonella* supernatant or derived from 5-A-RU/methylglyoxal condensation (data not shown, Extended Data Fig. 5a). We therefore evaluated whether the respective MR1 tetramers formed from these distinct synthetic antigens were similar functionally. MR1-5-OP-RU and MR1-5-OE-RU tetramers efficiently stained all human MAIT cells present in peripheral blood mononuclear cells (PBMCs), similarly to the mutant MR1(K43A) tetramers[6] (Fig. 4a, b). We solved the crystal structure of the MAIT TCR–MR1(K43A)–antigen complex, which revealed 5-OP-RU as the ligand bound to MR1(K43A), indicating that MR1 captures an intermediate from the synthesis of rRL-6-CH_2OH (Extended Data Fig. 4 and Extended Data Table 1). Thus, active MAIT-cell ligands are intermediary, open-ring precursors to ribityllumazines that arise from condensing 5-A-RU with small molecule metabolites.

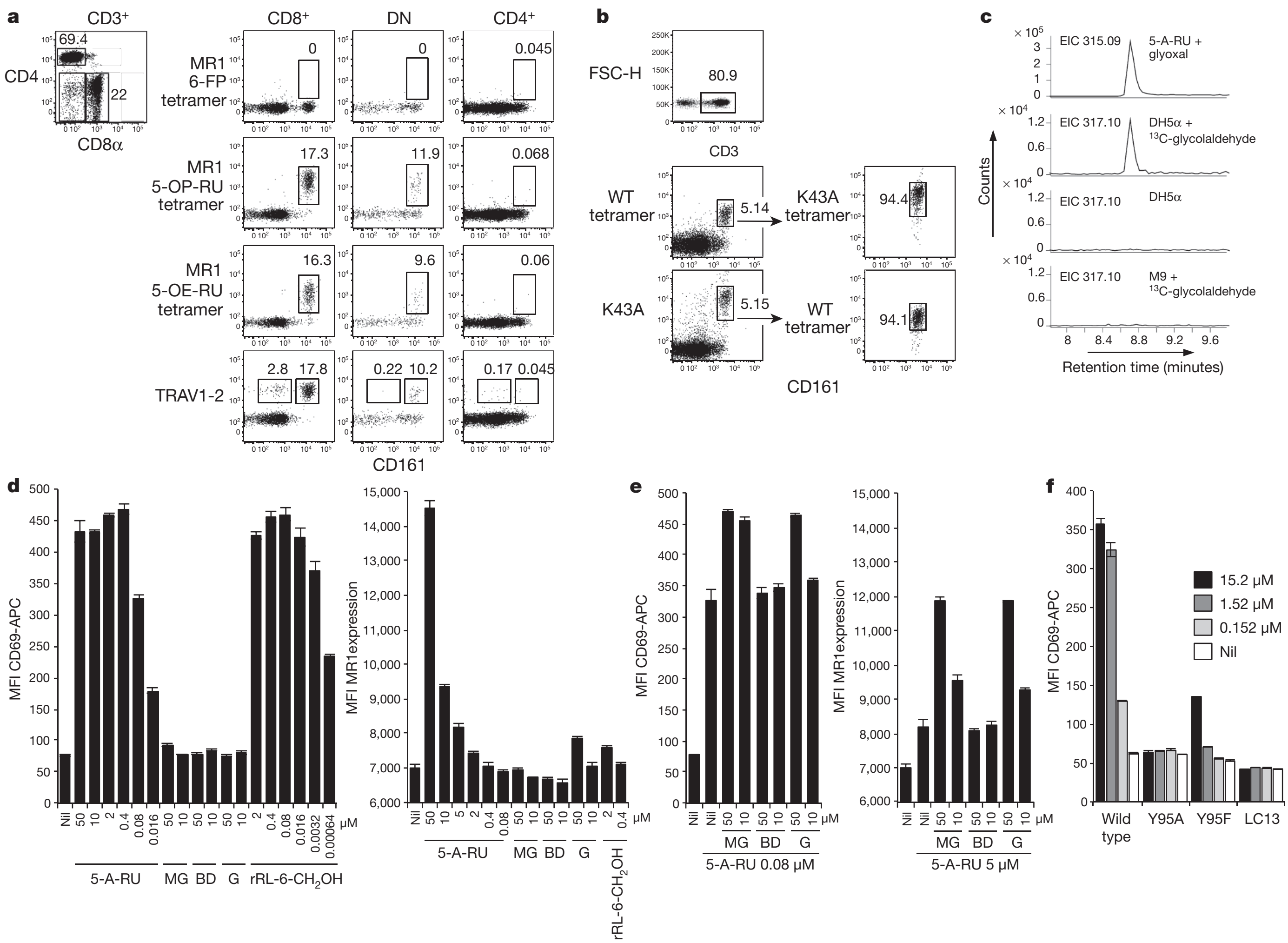


Figure 4 | MR1–antigen tetramers and MAIT activation. **a**, Gating strategy (left), tetramers of MR1–6-FP, MR1–5-OP-RU, MR1–5-OE-RU or anti-TRAV1-2. DN, double negative. **b**, PBMC co-staining with MR1–5-OP-RU and MR1(K43A)–5-OP-RU tetramers. WT, wild type. **c**, Extracted ion chromatograms (EICs) of *m*/*z* 315.09 or *m*/*z* 317.10. **d**, Left, activation assay and, right, MR1 upregulation with 5-A-RU, methylglyoxal (MG), butane-2,3-dione (BD), glyoxal (G) or rRL-6-CH_2OH. MFI of monoclonal antibody 26.5-PE staining. **e**, Left, CD69 upregulation and, right, MR1 upregulation with 5-A-RU, MG, BD or G. MFI of 26.5-PE. **f**, Activation of wild-type, Y95A- or Y95F-mutant SKW.MAIT cells by 5-A-RU. Data are shown as mean of triplicates plus s.e.m. Experiments were performed at least twice (**b**, **c**, **f**) or three times (**a**, **d**, **e**).

Recombinant MR1 refolded in the presence of folate-deficient culture supernatant from *S. typhimurium* SL1344 captured a dominant species of *m*/*z* 329.11 (ref. 5). Mass spectrometry of MR1 refolded with supernatant from *E. coli* (DH5α) also revealed a distinct and abundant *m*/*z* 315.09 species with matching liquid chromatography retention time, mass spectrometry and tandem mass spectrometry (MS/MS) properties to those observed with MR1-5-OE-RU (Extended Data Fig. 6 and data not shown). Closer analysis of the MR1 eluate from *S. typhimurium* also revealed the presence of an *m*/*z* 315.09 species, although this ligand was much less prevalent (data not shown). Accordingly, bacteria with an active riboflavin pathway can produce distinct MAIT-activating ligands, the relative abundance of which is dependent upon the bacterial source.

Next, to establish if bacteria also produced free 5-A-RU, ^{13}C-labelled glycolaldehyde was added to *E. coli* supernatant, which was subsequently refolded with MR1. We detected a species with an *m*/*z* of 317.1 from these MR1 eluates (Fig. 4c), consistent with the *m*/*z* 317.1 species observed previously (Extended Data Fig. 5). This indicates that there is sufficient free 5-A-RU released by bacteria to conjugate with exogenously added metabolites. Potent MAIT-cell antigens could also potentially be generated by host-derived metabolites forming adducts with 5-A-RU, in a manner somewhat analogous to the genesis of a CD1b-restricted antigen[17]. To test this, we added 5-A-RU to C1R cells transduced with MR1 (C1R.MR1), which led to MR1 cell surface upregulation and activation of Jurkat.MAIT cells (Fig. 4d, e). When exogenous glyoxal or methylglyoxal were added with 5-A-RU to C1R.MR1 cells, we observed a further increase in MR1 upregulation and an increase in Jurkat.MAIT activation, when compared with 5-A-RU added by itself (Fig. 4d, e). Notably, MR1 surface expression was not enhanced, nor was there an increase in Jurkat.MAIT activation, upon co-addition of butane-2,3-dione with 5-A-RU (Fig. 4d, e). These observations suggest that MR1–antigen complexes created from 5-A-RU and glyoxal or methylglyoxal are natively conformed. To test this, the mutation of Tyr 95α to either Ala 95 or Phe 95 ablated recognition of C1R cells to which 5-A-RU had been added, in a manner similar to that observed when synthetic rRL-6-CH_2OH was added to C1R cells[16], consistent with the notion that 5-A-RU is converted to 5-OE-RU or 5-OP-RU within C1R antigen-presenting cells (Fig. 4f). Accordingly, the bacterial riboflavin metabolite 5-A-RU can interact with host-derived metabolites analogous functionally to the creation of MR1 ligands found in bacterial supernatant.

We show a unique mechanism for creating T-cell ligands from disparate metabolite building blocks. The potent MAIT-activating ligands arise from a 'core precursor element' of the microbial riboflavin pathway that forms simple adducts with distinct chemical metabolites, via a mechanism that does not require enzymatic catalysis. Thus, MR1 captures, stabilizes and presents otherwise transitory chemical intermediates for MAIT-cell recognition. This represents a sophisticated discriminatory mechanism for targeting microbial antigens and protecting the host, whereby distinct metabolic pathways converge to produce T-cell antigens.

METHODS SUMMARY

L. lactis strains NZ9000 (wild type), the NZ9000 RibA^{-} deletion mutant, and the CB013 and CB021 roseoflavin resistant mutants have been previously described[11]. *S. typhimurium* strains SL1344 and BRD509 have been previously described[18]. See Methods for full details of CB013 derivatives and *Salmonella* Δ*ribDH* mutants. For details on chemical syntheses, characterization and stability, see Methods, Extended Data Figs 2, 3, 7, 8 and Supplementary Information. MAIT-activation and MR1-upregulations assays, MR1 refolding, and mass spectrometry analysis were conducted as described previously[5]. MR1(K43A) tetramers were generated as described[6]. See Methods for MR1–5-OP-RU and MR1–5-OE-RU tetramers. MR1 tetramer staining was conducted as described previously[6]. The MAIT TCR in complex with MR1–5-OP-RU, MR1–5-OE-RU and MR1(K43A)–5-OP-RU were crystallized and the structures were determined, as described in Methods.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 26 August 2013; accepted 14 February 2014.
Published online 2 April 2014.

1. Brigl, M. & Brenner, M. B. CD1: antigen presentation and T cell function. *Annu. Rev. Immunol.* **22,** 817–890 (2004).
2. Yewdell, J. W. & Haeryfar, S. M. Understanding presentation of viral antigens to CD8^{+} T cells *in vivo*: the key to rational vaccine design. *Annu. Rev. Immunol.* **23,** 651–682 (2005).
3. Eckle, S. B. G., Turner, S. J., Rossjohn, J. & McCluskey, J. Predisposed αβ T cell antigen receptor recognition of MHC and MHC-I like molecules? *Curr. Opin. Immunol.* **25,** 653–659 (2013).
4. Rossjohn, J., Pellicci, D. G., Patel, O., Gapin, L. & Godfrey, D. I. Recognition of CD1d-restricted antigens by natural killer T cells. *Nature Rev. Immunol.* **12,** 845–857 (2012).
5. Kjer-Nielsen, L. *et al.* MR1 presents microbial vitamin B metabolites to MAIT cells. *Nature* **491,** 717–723 (2012).
6. Reantragoon, R. *et al.* Antigen-loaded MR1 tetramers define T cell receptor heterogeneity in mucosal-associated invariant T cells. *J. Exp. Med.* **210,** 2305–2320 (2013).
7. Gold, M. C. & Lewinsohn, D. M. Co-dependents: MR1-restricted MAIT cells and their antimicrobial function. *Nature Rev. Microbiol.* **11,** 14–19 (2013).
8. Treiner, E. *et al.* Selection of evolutionarily conserved mucosal-associated invariant T cells by MR1. *Nature* **422,** 164–169 (2003).
9. Birkinshaw, R. W., Kjer-Nielsen, L., Eckle, S. B. G., McCluskey, J. & Rossjohn, J. MAITs, MR1 and vitamin B metabolites. *Curr. Opin. Immunol.* **26,** 7–13 (2014).
10. Vitreschak, A. G., Rodionov, D. A., Mironov, A. A. & Gelfand, M. S. Riboswitches: the oldest mechanism for the regulation of gene expression? *Trends Genet.* **20,** 44–50 (2004).
11. Burgess, C., O'Connell-Motherway, M., Sybesma, W., Hugenholtz, J. & van Sinderen, D. Riboflavin production in *Lactococcus lactis*: potential for *in situ* production of vitamin-enriched foods. *Appl. Environ. Microbiol.* **70,** 5769–5777 (2004).
12. Cushman, M. *et al.* Design, synthesis, and evaluation of 6-carboxyalkyl and 6-phosphonoxyalkyl derivatives of 7-oxo-8-ribitylaminolumazines as inhibitors of riboflavin synthase and lumazine synthase. *J. Org. Chem.* **67,** 5807–5816 (2002).
13. Bacher, A., Eberhardt, S., Fischer, M., Kis, K. & Richter, G. Biosynthesis of vitamin B2 (riboflavin). *Annu. Rev. Nutr.* **20,** 153–167 (2000).
14. Kis, K., Kugelbrey, K. & Bacher, A. Biosynthesis of riboflavin. The reaction catalyzed by 6,7-dimethyl-8-ribityllumazine synthase can proceed without enzymatic catalysis under physiological conditions. *J. Org. Chem.* **66,** 2555–2559 (2001).
15. Wang, Y. & Ho, C.-T. Flavour chemistry of methylglyoxal and glyoxal. *Chem. Soc. Rev.* **41,** 4140–4149 (2012).
16. Patel, O. *et al.* Recognition of vitamin B metabolites by mucosal-associated invariant T cells. *Nature Commun.* **4,** 2142 (2013).
17. Moody, D. B. *et al.* Cd1b-mediated T cell recognition of a glycolipid antigen generated from mycobacterial lipid and host carbohydrate during infection. *J. Exp. Med.* **192,** 965–976 (2000).
18. Hoiseth, S. K. & Stocker, B. A. D. Aromatic-dependent *Salmonella typhimurium* are non-virulent and effective as live vaccines. *Nature* **291,** 238–239 (1981).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the staff of the Monash crystallization facility and the Australian Synchrotron for assistance with crystallization and data collection, respectively, and M. J. McConville, P. O'Donnell and C.-S. Ang from the University of Melbourne Bio21 Institute mass spectrometry platform for assistance with mass spectrometry experiments. The Australian Research Council (ARC), a Program Grant and Project Grant of the National Health and Medical Research Council of Australia (NHMRC) supported this research; D.V.S. is a recipient of a Science Foundation Ireland (SFI) Principal Investigator award (ref. no. 08/IN.1/B1909); O.P. was supported by an ARC Future Fellowship; D.P.F. was supported by an NHMRC Senior Principal Research Fellowship (1027369); J.R. was supported by an NHMRC Australia Fellowship.

Author Contributions A.J.C., S.B.G.E., R.W.B. and L.L. are joint first authors. O.P., J.Ma., H.C., N.A.W., R.A.S., Z.C., R.R., B.M., D.V.S. and J.Y.W.M. either performed experiments, provided key reagents, and/or analysed data, and/or provided intellectual input or helped write the manuscript. D.P.F., L.K.N., J.Mc. and J.R. co-led the investigation and contributed to design and interpretation of data, project management, and writing of the manuscript. D.P.F., L.K.N., J.Mc. and J.R. are joint senior authors.

Author Information The atomic coordinates and structure factors for the TCR–MR1–antigen complexes have been deposited in the Protein Data Bank under accession numbers 4NQC, 4NQD and 4NQE. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.P.F. (d.fairlie@imb.uq.edu.au), L.K.N. (lkn@unimelb.edu.au), J.Mc. (jamesm1@unimelb.edu.au) or J.R. (jamie.rossjohn@monash.edu).

METHODS

Bacterial strains and mutants. *L. lactis* strains NZ9000 (wild type), the NZ9000 $RibA^-$ deletion mutant, and the CB013 and CB021 roseoflavin resistant mutants have been previously described[11]. The CB013 derivatives CB013ΔRibA, CB013ΔRibB, CB013ΔRibG and CB013ΔRibH were generated by insertion of EcoRI or EcoRV restriction sites incorporating either one or two stop codons into the individual genes using standard techniques. Inserted sequences were as follows (indicated in uppercase). ΔRibG, attaacgtttccccctcctttcgagccagtGAATTCaggattgctaaattcataaa atgctcatcattttccat; ΔRibB, gatgctggaagctcgaatgattaatttagacgGAATTCATTAacttat ctctcttttgaatttgagttacctctccttat; ΔRibA tatcttttctggactaatcatttcggctgcacGAATTCaa atcagatctccttcattttctcttattctcatcatc, ΔRibH, ggcccctgctaagagttttgcgttgatgaaGATAT CTTATTAttcgttaaaacgtgcaactacaattccatatttgg.

The genotype of each mutant was verified by sequencing and multiple rounds of PCR based on the mutated region to verify the purity of the genotype. The phenotype of each mutant was also checked using growth or absence of growth in riboflavin assay medium overnight at 30 °C and comparing with a control of CB013. All mutants were unable to grow in the media whereas CB013 was capable of growth in riboflavin-limiting conditions as it is an over-producer. *L. lactis* strains were grown at 30 °C without shaking in M17 medium (Difco) containing 1% glucose and the addition of 3 $\mu g\,ml^{-1}$ riboflavin where indicated. *S. typhimurium* SL1344 was grown at 37 °C without shaking in M17 medium (Difco) containing 1% glucose.

S. typhimurium strains SL1344 and BRD509 have been previously described[18]. The *Salmonella* Δ*ribDH* mutants were constructed on an SL1344 background by lamda-red-recombinase-mediated allelic replacement followed by general transduction using phage P22 as previously described[19], resulting in strain SL1344ΔRibDH.

Primers were as follows. B2(Sec)F, 5′-TAGGGATAACAGGGTAAT-GGTTC GATAGCGTAATGG; B2(Sec)R, 5′-TAGGGATAACAGGGTAAT-TATCTTTC CGGCCTGTGA; B2(Kan)F, 5′-CTAAGGAGGATATTCATATG-GACCGCGC TTGAAATGAT; B2(Kan)R, 5′-GAAGCAGCTCCAGCCTACACA-ATTGTTA ACAATGACACA.

The complement of mutants was performed by transformation of *ribDH* genes, resulting in strain SL1344ΔRibDH:RibDH. Mutation and reconstitution were verified by lack of growth or growth on Luria agar, and by PCR. Mutants were grown on Luria agar containing 20 $\mu g\,ml^{-1}$ riboflavin.

For MR1 refolds, *Salmonella* wild-type and mutant strains were grown in M9 minimal media supplemented with histidine (77.6 $\mu g\,ml^{-1}$) and streptomycin (25 $\mu g\,ml^{-1}$) and 3 $\mu g\,ml^{-1}$ riboflavin. *E. faecalis* was grown in Folic Acid Assay Medium (Difco) at 37 °C without agitation. *E. coli* DH5α was grown in M9 media. *L. lactis* CB013 and CB013 riboflavin mutants were grown in Folic Acid Assay Medium (Difco) supplemented with xanthine (6 $\mu g\,ml^{-1}$) and Yeast Nitrogen Base (6.8 $mg\,ml^{-1}$) at 30 °C without agitation.

Compounds. Glyoxal, methylglyoxal, 1,3-dihydroxyacetone dimer, DL-glyceraldehyde and butane-2,3-dione were purchased from Sigma. [1,2-$^{13}C_2$]glycolaldehyde (glycolaldehyde can be readily air-oxidized to form glyoxal) was purchased from Omicron Biochemicals. A synthesis of rRL-6-CH_2OH has been previously described[5]. 5-A-RU was freshly prepared from 5-nitroso-6-D-ribitylaminouracil following a previously described procedure[20]. In brief, 5-nitroso-6-D-ribitylaminouracil (40.0 mg, 0.138 mmol, 1 eq) was dissolved in MilliQ water (3 ml) at 80 °C under argon. To the red solution was added sodium dithionite powder (1.2–3.3 eq). The colour changed instantly to pale yellow. After stirring at 80 °C for 5 min, the solution was cooled under argon in an ice-water bath. For biological studies, the chilled solution was diluted with MilliQ water to make 50 mM stock solutions and stored in 1.5 ml aliquots at −20 °C for later use.

For NMR characterization of the pyrimidine intermediate 5-OP-RU (**3d**), a freshly prepared solution of 5-A-RU (**1**) was adjusted to pH 7.0 with 1 M sodium hydroxide solution, lyophilized, dissolved in DMSO-d_6 and then filtered to remove salts. The solution was transferred to an NMR tube, filled with argon, and the concentration of 5-A-RU was determined by NMR spectroscopy. Methylglyoxal (2 eq) was added, and the reaction monitored by NMR. Upon completion, 5-OP-RU was further purified using a Shimadzu preparative HPLC system equipped with a Phenomenex Luna 10 μm C18 250 × 21.20 mm column (P/No 00G-4253-PO-AX) and a SPD-M20A diode array detector. Flow rate was 20 $ml\,min^{-1}$ with linear gradient: 100% solvent A to 100% solvent B over 30 min, where solvent A was 20 mM ammonium acetate in H_2O and solvent B was 20 mM ammonium acetate in MeCN–H_2O (80:20, v/v). Compound **3d** was fully characterized by ESI-HRMS (Calculated for $C_{12}H_{17}N_4O_7^-$ *m/z* 329.1103, measured *m/z* 329.1116) and one-dimensional and two-dimensional NMR spectroscopy (Fig. 2b and Extended Data Fig. 2).

5-OP-RU (**3d**) ^1H NMR (600 MHz, DMSO-d_6), δ 2.28 (3H, s), 3.38–3.43 (2H, m), 3.47–3.51 (1H, m), 3.52–3.55 (1H, m), 3.56–3.59 (2H, m), 3.73 (1H, m), 7.43 (1H, br s), 8.80 (1H, s); ^{13}C NMR (150 MHz, DMSO-d_6) δ 23.5, 44.1, 63.1, 70.7, 72.8, 72.9, 98.5, 142.0, 152.1, 157.6, 159.1, 200.2. ESI-HRMS calculated for $C_{12}H_{17}N_4O_7^-$ $[M-H]^-$: 329.1103, found: 329.1116.

Stability of 5-OP-RU in aqueous media. Purified 5-OP-RU was dissolved in aqueous TBS buffer (10 mM Tris, 150 mM NaCl, pH 8.0), MilliQ water (pH 6.8), or aqueous ammonium acetate buffer (20 mM, pH 5.4). The consumption of 5-OP-RU was immediately monitored by LC–MS. The initial concentration was quantified by comparing with a standard solution of known concentration. At 15 °C, the half-life was determined as 14.5–15 h at pH 8.0 independent of the starting concentrations (65–250 μM), 14.2 h at pH 6.8 (65 μM), and 49 min at pH 5.4 (65 μM). At 37 °C, pH 6.8, the half-life was 135 min (Extended Data Fig. 3).

Activation of Jurkat.MAIT and SKW.MAIT cells and detection of MR1 expression on C1R.MR1 cells. Jurkat cells transduced with genes encoding a MAIT TCR comprising the TRAV1-2–TRAJ33 invariant α chain, and a TRBV6-1 β chain, or SKW cells transduced with genes encoding the TRAV1-2–TRAJ33 invariant α chain with either wild-type Tyr 95 or mutated Tyr95Ala or Tyr95Phe residues, paired with a TRBV6-1 β chain, were tested for activation by co-incubation with bacterial culture supernatant or compounds and C1R antigen presenting cells expressing MR1 (CIR.MR1, with Jurkat.MAIT cells), or C1R cells (SKW.MAIT cells) for 16 h. Cells were subsequently stained with PE-Cy7-conjugated anti-CD3 (eBioscience), and APC-conjugated anti-CD69 (BD) antibodies as well as biotinylated anti-MR1 monoclonal antibody 26.5 (ref. 21), followed by Streptavidin-PE (BD), before analysis by flow cytometry. Activation of Jurkat.MAIT or SKW. MAIT cells was measured by an increase in surface CD69 expression. MR1 expression was detected on gated C1R.MR1 cells in the same assay.

Preparation of denatured inclusion body MR1 and β_2M. Genes encoding soluble human MR1 or human β_2M were expressed for 4 h in BL21 *E. coli* after induction with 1 mM isopropyl β-D-1-thiogalactopyranoside. *E. coli* were pelleted and resuspended in a buffer containing 50 mM Tris, 25% (w/v) sucrose, 1 mM EDTA, 10 mM dithiothreitol (DTT) pH 8.0. Inclusion body protein was then extracted by lysis of bacteria in a buffer containing 50 mM Tris pH 8.0, 1% (w/v) Triton X-100, 1% (w/v) sodium deoxycholate, 100 mM NaCl, 10 mM DTT, 5 mM $MgCl_2$ and 1 mg DNaseI per litre of starting culture; and subsequent steps involved homogenization with a polytron homogenizer, centrifugation and washing inclusion body protein sequentially with first a buffer containing 50 mM Tris pH 8.0, 0.5% Triton X-100, 100 mM NaCl, 1 mM EDTA, 1 mM DTT, and second a buffer containing 50 mM Tris pH 8.0, 1 mM EDTA, 1 mM DTT. Inclusion body protein was then resuspended in a buffer containing 20 mM Tris pH 8.0, 8 M urea, 0.5 mM EDTA, 1 mM DTT, and following centrifugation the supernatant containing solubilized, denatured inclusion body protein was collected and stored at −80 °C.

Refolding of MR1 ligand and MAIT TCR. As MR1 cannot be expressed in a soluble form in the absence of ligand, MR1 (the ectodomain) and β_2M were refolded with ligand essentially as described[5]. Briefly, in order to generate MR1–5-OP-RU and MR1–5-OE-RU, 56 mg of MR1 and 28 mg of β_2M inclusion body proteins, together with 2.9 mg of 5-A-RU and 254, 204 or 204 mg of methylglyoxal, glyoxal (Sigma) or ^{13}C-glycolaldehyde (Omicron Biochemicals), respectively, were added to a 400 ml refold solution containing 0.1 M Tris, pH 8.5, 2 mM EDTA, 0.4 M arginine, 0.5 mM oxidized glutathione and 5 mM reduced glutathione. Refolded MR1 antigen was then purified by sequential DEAE (GE Healthcare) anion exchange, S75 16/60 (GE Healthcare) gel filtration, and MonoQ (GE Healthcare) anion exchange chromatography. Alternatively, 56 mg of MR1 and 28 mg of β_2M inclusion body proteins were refolded in the presence of 400 ml of 0.45 μM filtered bacterial supernatants or control media, in the absence or presence of 204 mg ^{13}C-glycolaldehyde. The TRBV6-1 MAIT TCR (the ectodomains) was expressed, refolded and purified essentially as previously described[22].

Sequences of constructs used in refolding are as follows. Soluble TRAV1-2 α-chain amino acid sequence, excluding the transmembrane/cytoplasmic domains: MGQ NIDQPTEMTATEGAIVQINCTYQTSGFNGLFWYQQHAGEAPTFLSYNVLD GLEEKGRFSSFLSRSKGYSYLLLKELQMKDSASYLCAVKDSNYQLIWGAGTK LIIKPDIQNPDPAVYQLRDSKSSDKSVCLFTDFDSQTNVSQSKDSDVYITDKC VLDMRSMDFKSNSAVAWSNKSDFACANAFNNSIIPEDTFFPSPESS.

Soluble TRBV6-1 β-chain amino acid sequence, excluding the transmembrane/cytoplasmic domains: MNAGVTQTPKFQVLKTGQSMTLQCAQDMNHNSMY WYRQDPGMGLRLIYYSASEGTTDKGEVPNGYNVSRLNKREFSLRLESAAPS QTSVYFCASSVWTGEGSGELFFGEGSRLTVLEDLKNVFPPEVAVFEPSEAEIS HTQKATLVCLATGFYPDHVELSWWVNGKEVHSGVCTDPQPLKEQPALN DSRYALSSRLRVSATFWQNPRNHFRCQVQFYGLSENDEWTQDRAKPVTQ IVSAEAWGRAD. Soluble human MR1 amino acid sequence, excluding the transmembrane/cytoplasmic domains: MRTHSLRYFRLGVSDPIHGVPEFISVGYVD SHPITTYDSVTRQKEPRAPWMAENLAPDHWERYTQLLRGWQQMFKVEL KRLQRHYNHSGSHTYQRMIGCELLEDGSTTGFLQYAYDGQDFLIFNKDTL SWLAVDNVAHTIKQAWEANQHELLYQKNWLEEECIAWLKRFLEYGKDT LQRTEPPLVRVNRKETFPGVTALFCKAHGFYPPEIYMTWMKNGEEIVQEI DYGDILPSGDGTYQAWASIELDPQSSNLYSCHVEHSGVHMVLQVP.

Analysis of MR1-5-OP-RU and MR1-5-OE-RU by mass spectrometry. MR1-5-OP-RU or MR1-5-OE-RU (4 μg) were loaded onto an XBridge C18 reversed phase

column (Waters) in 20 mM ammonium acetate, pH 5.4, buffer, and detected in an Agilent ESI–TOF mass spectrometer after elution in an acetonitrile gradient. Data were collected in negative ion mode. Different instrumentation resulted in slight variations in retention times of the *m/z* 329.11, 315.09 and 317.10 species to those reported previously[5].

Generation of MR1(K43A), MR1-5-OP-RU and MR1-5-OE-RU tetramers. The generation of MR1(K43A) tetramers, loaded with synthetic rRL-6-CH_2OH, has been previously described[6]. Briefly, refolded and purified empty carboxy-terminal cysteine-tagged-MR1(K43A) was loaded with a 136 molar excess of synthetic rRL-6-CH_2OH for 4 h at room temperature in the dark. C-terminal cysteine-tagged wild-type MR1-5-OP-RU and MR1-5-OE-RU were generated as described earlier.

Cysteine-tagged MR1(K43A)–5-OP-RU, or cysteine-tagged wild-type MR1–5-OP-RU or cysteine-tagged wild-type MR1–5-OE-RU were then reduced with 5 mM DTT for 20 min before buffer exchange into PBS using a PD-10 column (GE Healthcare), and biotinylated with Maleimide-PEG2 biotin (Thermoscientific) with a 30:1 molar ratio of biotin:protein at 4 °C for 16 h in the dark. Biotinylated MR1 was subjected to S200 10/300 GL (GE Healthcare) chromatography to remove excess biotin. Biotinylated, loaded MR1(K43A), or wild-type MR1–5-OP-RU or wild-type MR1–5-OE-RU monomers were tetramerized with streptavidin conjugated to either PE (SA-PE) or Brilliant Violet 421 (SA-BV) (BD Pharmingen).

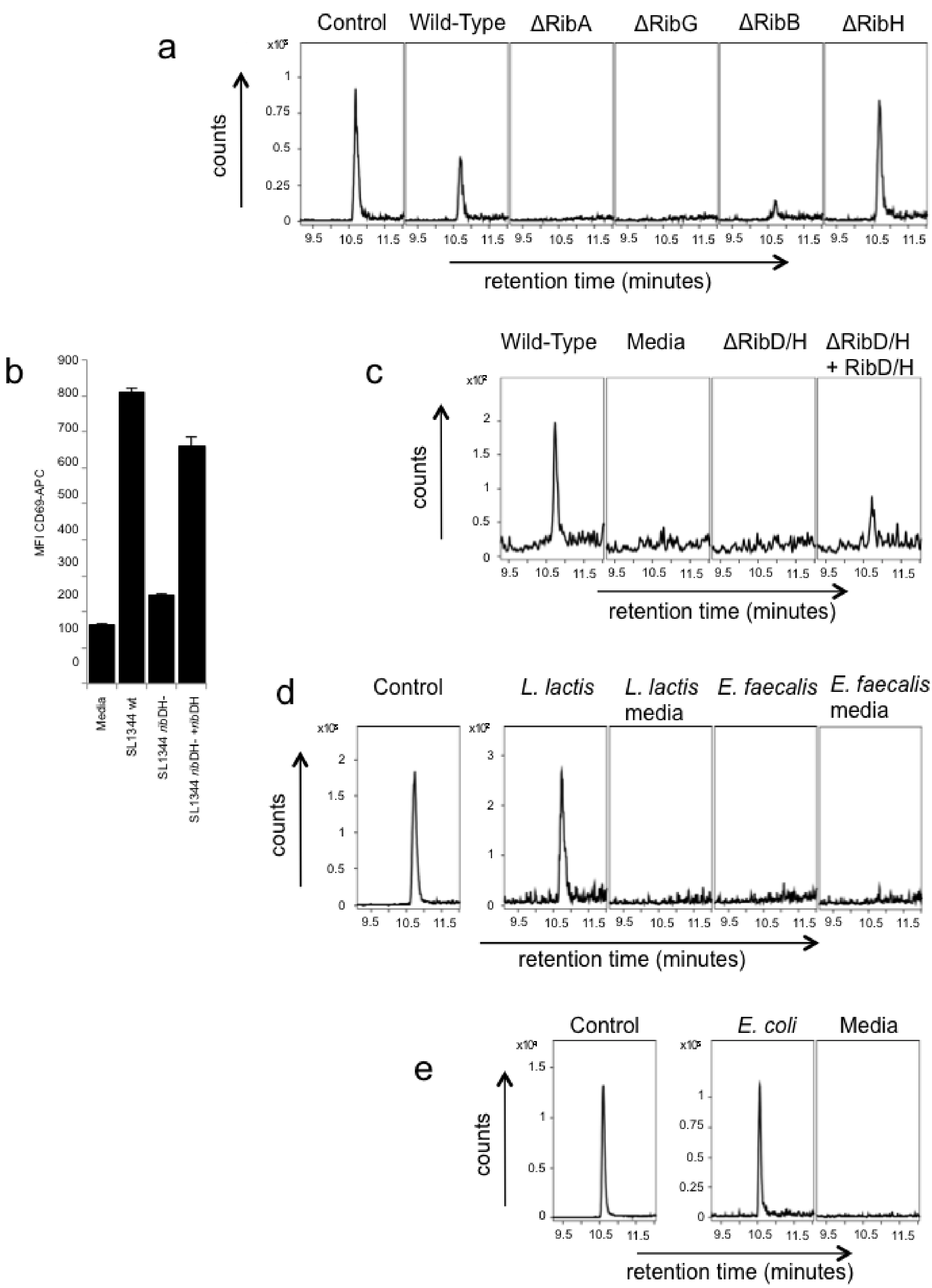
Isolation of PBMCs. Whole blood from healthy donors was collected (Australian Red Blood Cross Service) and PBMCs were separated using Ficoll-Paque Premium (GE Healthcare). PBMCs were harvested and resuspended in fresh RPMI medium. Cells were then washed twice before resuspension in 10% DMSO in FCS. Before use, PBMCs were stored in liquid nitrogen.

Tetramer staining of human PBMCs. For co-staining with wild-type and MR1 (K43A) tetramers, approximately 5×10^5 human PBMCs were stained with MR1 (K43A)–5-OP-RU-PE tetramer at 20 μg ml^{-1} for 40 min at room temperature in the dark. Cells were then washed and stained with wild-type MR1–5-OP-RU-BV tetramer at 1.4 μg ml^{-1}, CD3-AlexaFluor700 (EBioscience), CD161-PE-Cy7 (Biolegend), CD4-APC-Cy7 (Biolegend) and CD8α-PerCP-Cy5.5 (BD) for 30 min at 4 °C. Cells were then washed once with 2 ml of FACS wash (2% fetal bovine serum in PBS) and resuspended in 150 μl of FACS fix (2% glucose and 1% paraformaldehyde in PBS) before acquisition of data on a BD LSR-Fortessa. Data were analysed using FlowJo analysis software (Tree Star).

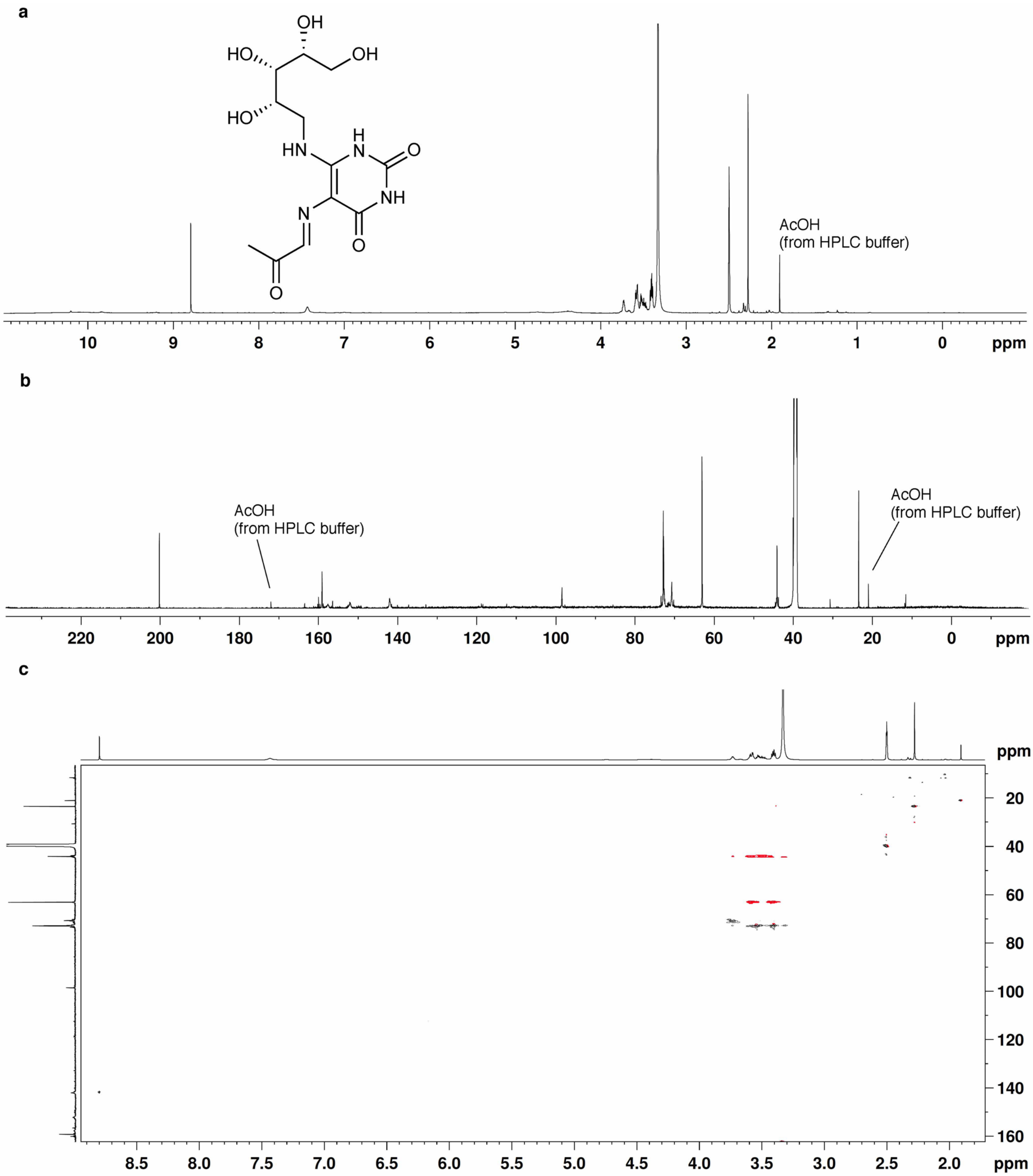
For single staining with either MR1–5-OP-RU or MR1–5-OE-RU tetramers, human PBMCs were stained as described earlier with wild-type MR1–5-OP-RU-PE or wild-type MR1–5-OE-RU-PE tetramers at 1.4 μg ml^{-1}, and CD3-AlexaFluor700 (EBioscience), CD161-PE-Cy7 (Biolegend), CD4-APC-Cy7 (Biolegend) and CD8α-PerCP-Cy5.5 (BD) for 30 min at 4 °C, before acquisition of data on a BD LSR-Fortessa.

Crystallization and structure determination. Crystals of the soluble MAIT TCR–MR1–antigen complexes were obtained using the hanging-drop vapour diffusion method. The MR1–β_2M–5-OP-RU, MR1–β_2M–5-OE-RU, MR1, MR1(K43A)–β_2M–5-OP-RU and MAIT TCR were concentrated to 4 mg ml^{-1}, mixed in a 1:1 molar ratio, then 0.5 μl added to 0.5 μl of a precipitant solution consisting 0.1 M Bis-Tris propane pH 6.3, 0.2 M sodium acetate and varying concentrations of PEG 3350 between 8–14% w/v. Crystals were observed after incubation at 20 °C for 24 h in dark conditions and cryoprotected before diffraction experiments by soaking in the crystallization condition modified with between 10–15% v/v glycerol before cooling to 100 K. Diffraction images were collected at the Australian Synchrotron MX2 beamline diffracting in a *C*2 space group to 2.50 Å, 2.10 Å and 2.20 Å for the MR1–β_2M–5-OP-RU, MR1–β_2M–5-OE-RU and MR1(K43A)–β_2M–5-OP-RU complexes with the MAIT TCR, respectively. The data were processed using Mosflm version 7.0.9 and scaled using AIMLESS or SCALA (MR1(K43A)–β_2M–5-OP-RU only) from the CCP4 suite[23]. The phase problem was solved by molecular replacement using PHASER[24], using MR1 ternary complex (PDB accession 4L4T)[16] with CDR loops and ligands removed and using the R_{free} reflection set from the model. The initial solution was refined in Phenix using simulated annealing refinement[25], with all subsequent refinement steps performed using BUSTER 2.10 (ref. 26). Restraints for 5-OP-RU and 5-OE-RU were generated using the Grade Web Server, with model building performed in COOT using MolProbity for validation[27]. All molecular graphics were made with PyMOL.

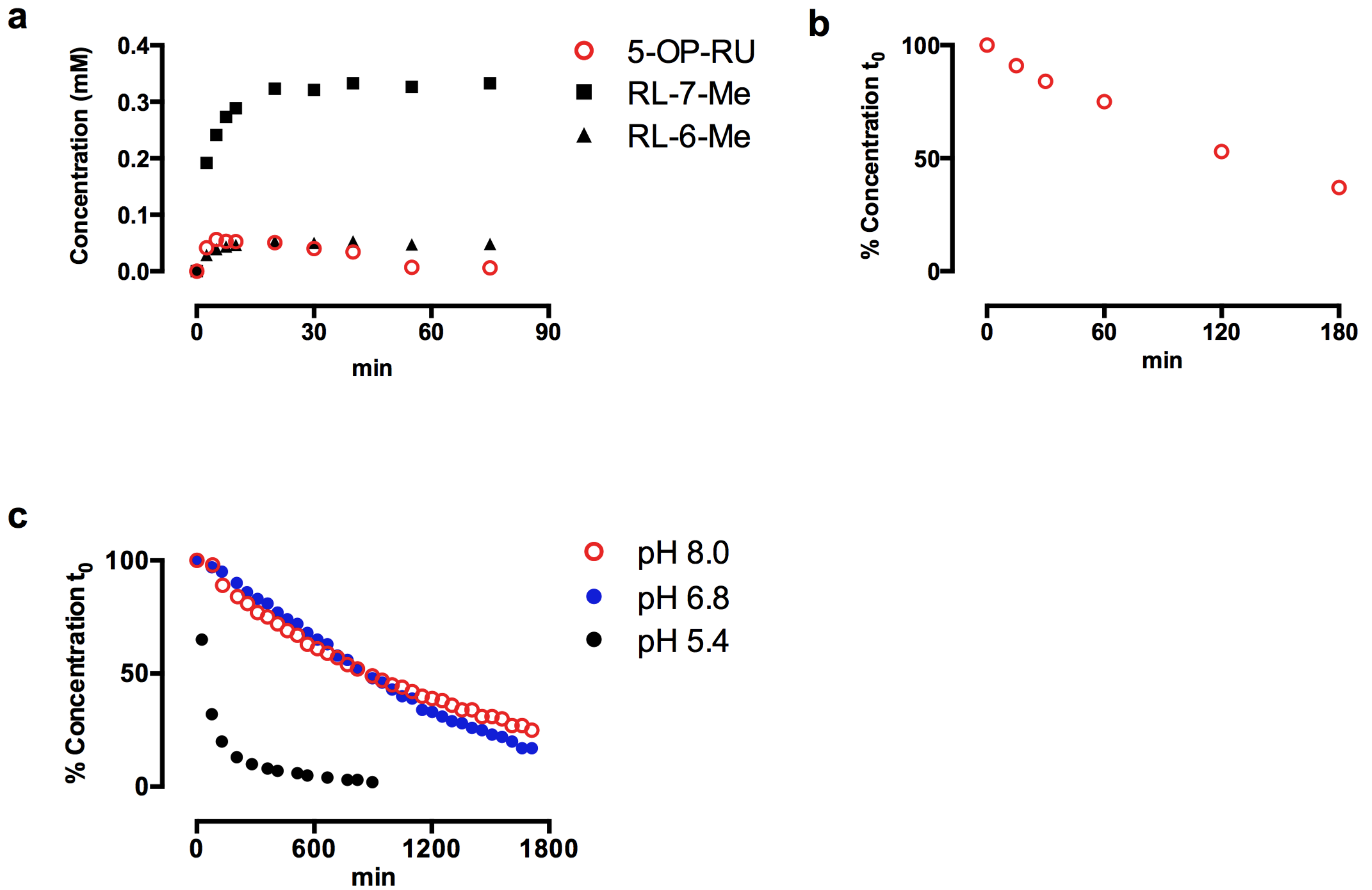
19. Strugnell, R. *et al.* Characterization of a *Salmonella typhimurium* aro vaccine strain expressing the P.69 antigen of *Bordetella pertussis*. *Infect. Immun.* **60,** 3994–4002 (1992).
20. Plaut, G. W. E. & Harvey, R. A. in *Methods in Enzymology* Vol. 18, Part B (eds McCormick, D. B. & Wright, L. D.) 515–538 (Academic, 1971).
21. Huang, S. *et al.* Evidence for MR1 antigen presentation to mucosal-associated invariant T cells. *J. Biol. Chem.* **280,** 21183–21193 (2005).
22. Reantragoon, R. *et al.* Structural insight into MR1-mediated recognition of the mucosal associated invariant T cell receptor. *J. Exp. Med.* **209,** 761–774 (2012).
23. CCP4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50,** 760–763 (1994).
24. McCoy, A. J. Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr. D* **63,** 32–41 (2007).
25. Zwart, P. H. *et al.* Automated structure solution with the PHENIX suite. *Methods Mol. Biol.* **426,** 419–435 (2008).
26. Smart, O. S. *et al.* Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER. *Acta Crystallogr. D* **68,** 368–380 (2004).
27. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).



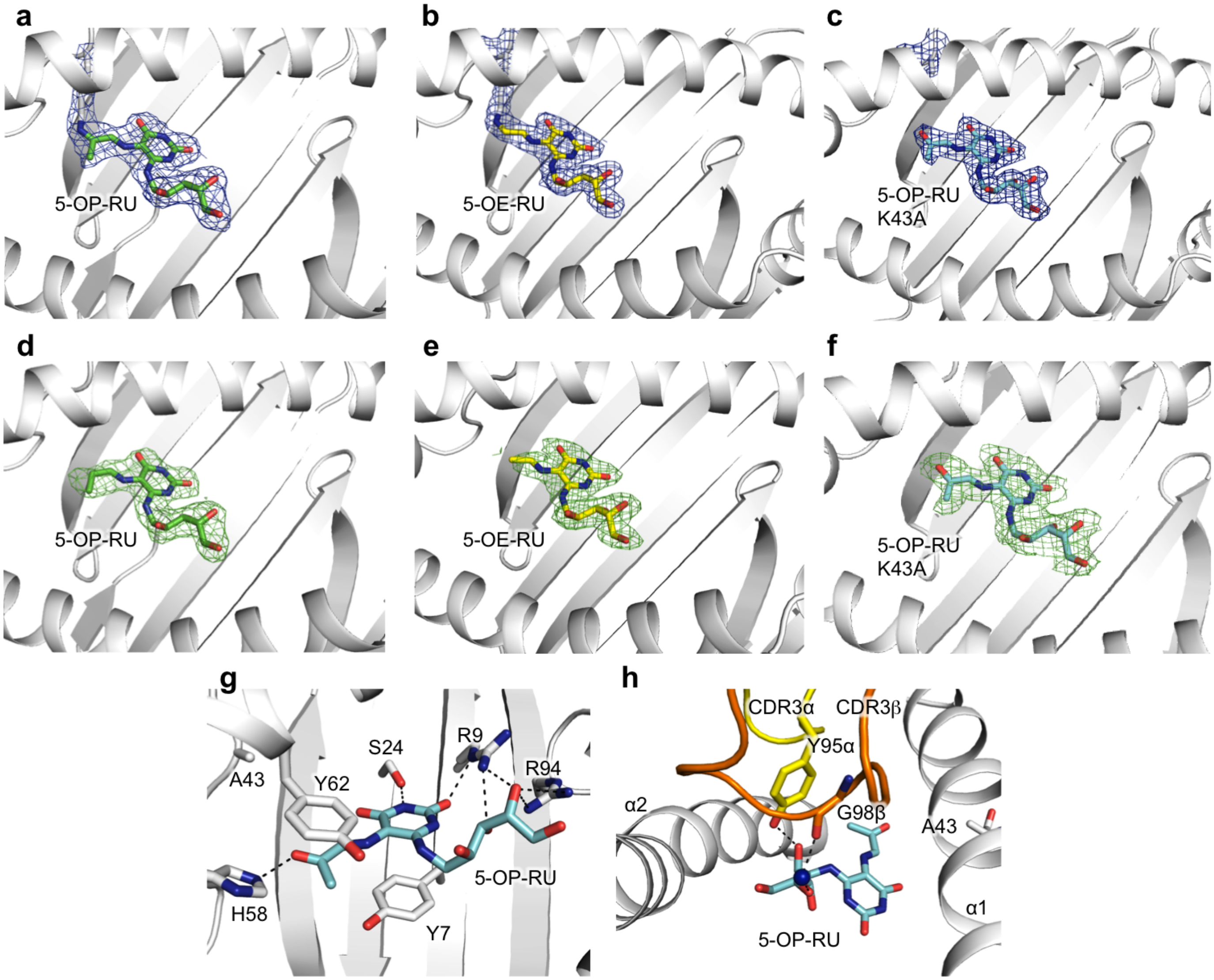
Extended Data Figure 1 | MR1 ligand identification from different bacterial strains. **a**, Detection of *m*/*z* 329.11 species in MR1 refolded with 5-A-RU and methylglyoxal (control), and supernatants from wild-type (CB013) and CB013-derivative (that is, ΔRibA, ΔRibB, ΔRibG or ΔRibH) *L. lactis* bacteria. Shown are counts on the *y*-axis versus retention time on the *x*-axis. **b**, Lack of activation of Jurkat.MAIT cells by supernatant from mutant ΔRibD/H *S. typhimurium* (strain SL1344) but not wild-type (WT), or ΔRibD/H plus RibD/H bacteria. Shown is MFI of CD69-APC on the *y*-axis. **c**, Detection of *m*/*z* 329.11 species in MR1 refolded with supernatants from wild-type, ΔRibD/H or ΔRibD/H plus RibD/H *S. typhimurium* bacteria, or control media. Shown are counts on the *y*-axis versus retention time on the *x*-axis. **d**, Detection of *m*/*z* 329.11 species in MR1 refolded with 5-A-RU and methylglyoxal (control), or bacterial supernatants from *L. lactis* (CB013) or *E. faecalis* bacteria, or control media. Shown are counts on the *y*-axis versus retention time on the *x*-axis. **e**, Detection of *m*/*z* 329.11 species in MR1 refolded with 5-A-RU and methylglyoxal (control), or supernatant from *E. coli* bacteria, or media. Shown are counts on the *y*-axis versus retention time on the *x*-axis. Experiments **a**–**e** were performed three, three, three, two and three times, respectively.



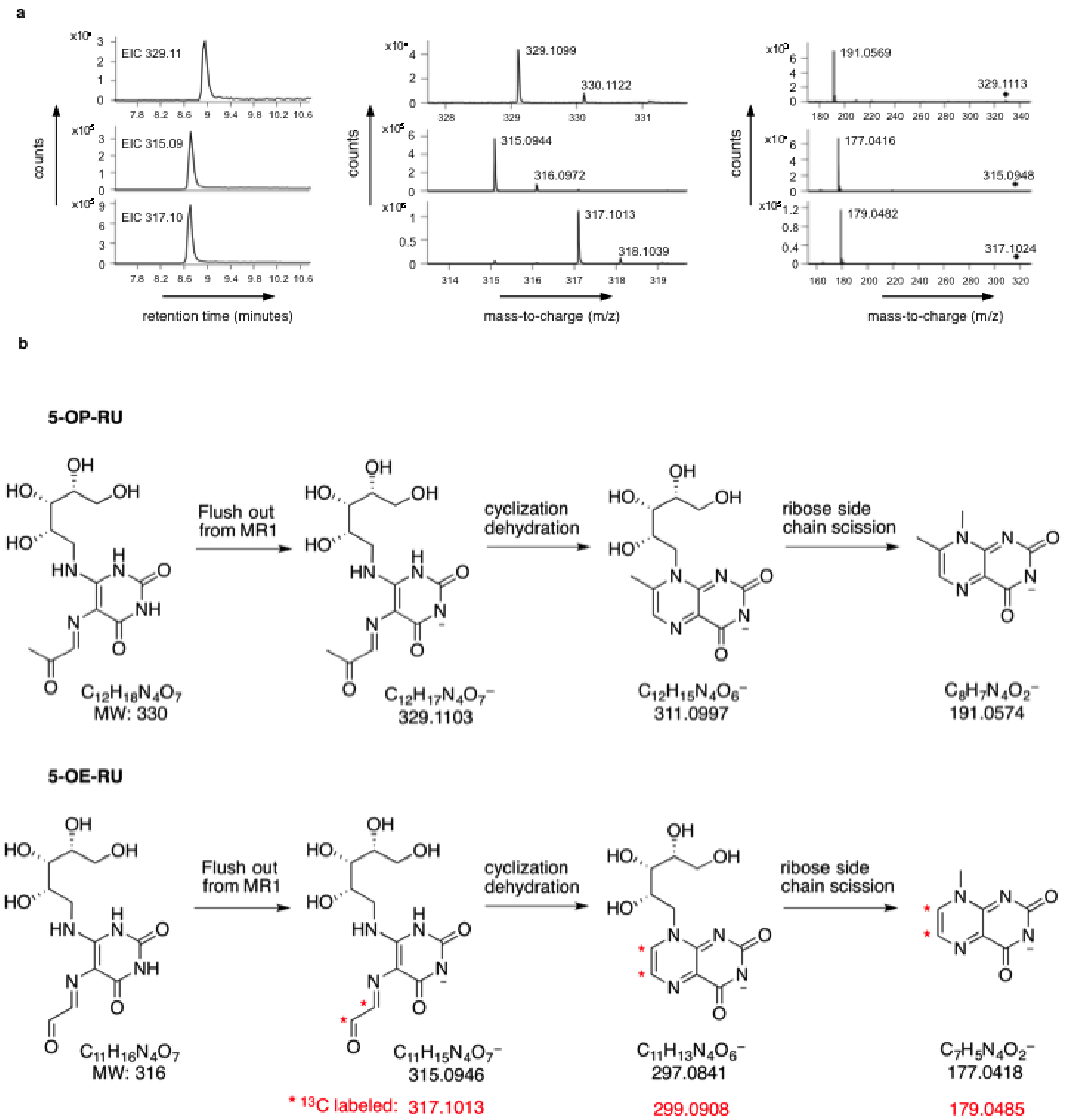
Extended Data Figure 2 | NMR characterization of 5-OP-RU. a–c, NMR characterization of 5-OP-RU (**3d**) in DMSO-d_6 with internal solvent peak at 2.50 p.p.m. and 39.52 p.p.m. for ^{1}H and ^{13}C, respectively. **a**, ^{1}H NMR (600 MHz); **b**, ^{13}C NMR (150 MHz); **c**, HSQC. The compound 5-OP-RU (**3d**) was synthesized from the reaction of 5-A-RU and methylglyoxal in DMSO-d_6, and then isolated from aqueous media by reversed-phase high-performance liquid chromatography (rpHPLC). Although it was less stable in water, it could still be identified and characterized at pH > 6.



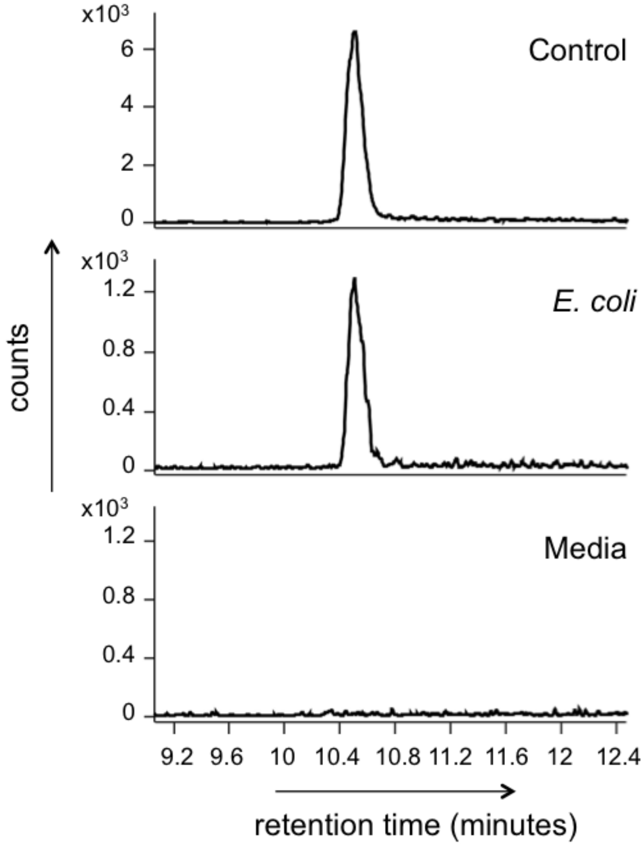
Extended Data Figure 3 | Stability of 5-OP-RU. **a**, Reaction between 5-A-RU (0.5 mM) and methylglyoxal (3 equivalents (eq)) at pH 6.8, 37 °C in MilliQ water. **b**, Stability of purified 5-OP-RU (65 μM) at pH 6.8 and 37 °C. The half-life was 135 min. **c**, Stability of purified 5-OP-RU (65 μM) at variable pH in aqueous TBS buffer (10 mM Tris, 150 mM NaCl, pH 8.0), MilliQ water (pH 6.8) or ammonium acetate buffer (20 mM, pH 5.4) at 15 °C. The half-lives were 15 h at pH 8.0, 14.2 h at pH 6.8, 49 min at pH 5.4.



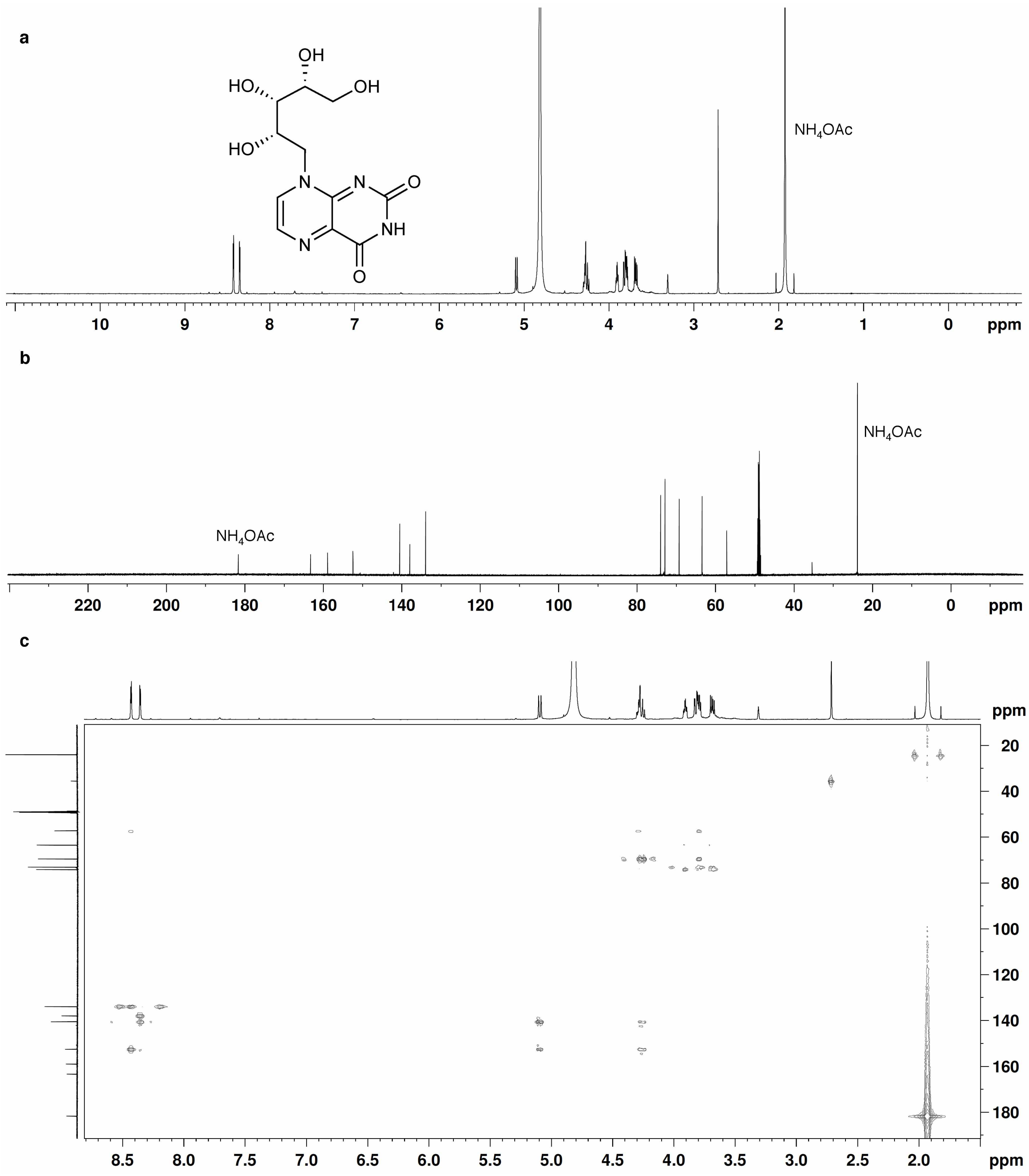
Extended Data Figure 4 | Electron density for ligands, and contacts associated with MR1(K43A)–5-OP-RU MAIT TCR complex. a–h, Electron density of 5-OP-RU in MR1, 5-OE-RU in MR1 and 5-OP-RU in MR1(K43A). **a–c**, Final $2F_o - F_c$ map, contoured at 1σ for 5-OP-RU (**a**) and 5-OE-RU (**b**) in the MAIT TCR–MR1–antigen complex, and 5-OP-RU (**c**) in the MAIT TCR–MR1(K43A)–antigen complex. **d–f**, Simulated annealing omit maps showing unbiased $F_o - F_c$ electron density, contoured at 3σ, for 5-OP-RU (**d**) and 5-OE-RU (**e**) in MR1, and 5-OP-RU (**f**) in MR1(K43A). **g**, **h**, MR1(K43A)–5-OP-RU MAIT TCR complex showing contacts between MR1(K43A) and 5-OP-RU (**g**) and contacts between MAIT TCR and 5-OP-RU (**h**). MR1 is shown in grey, MAIT TCR CDR3α in yellow and CDR3β in orange with ribbon representation, and 5-OP-RU is shown in cyan with stick representation. Hydrogen bonds are indicated with black dashed lines with a water molecule mediating hydrogen bonding between the CDR3β 5-OP-RU shown in dark blue sphere representation.



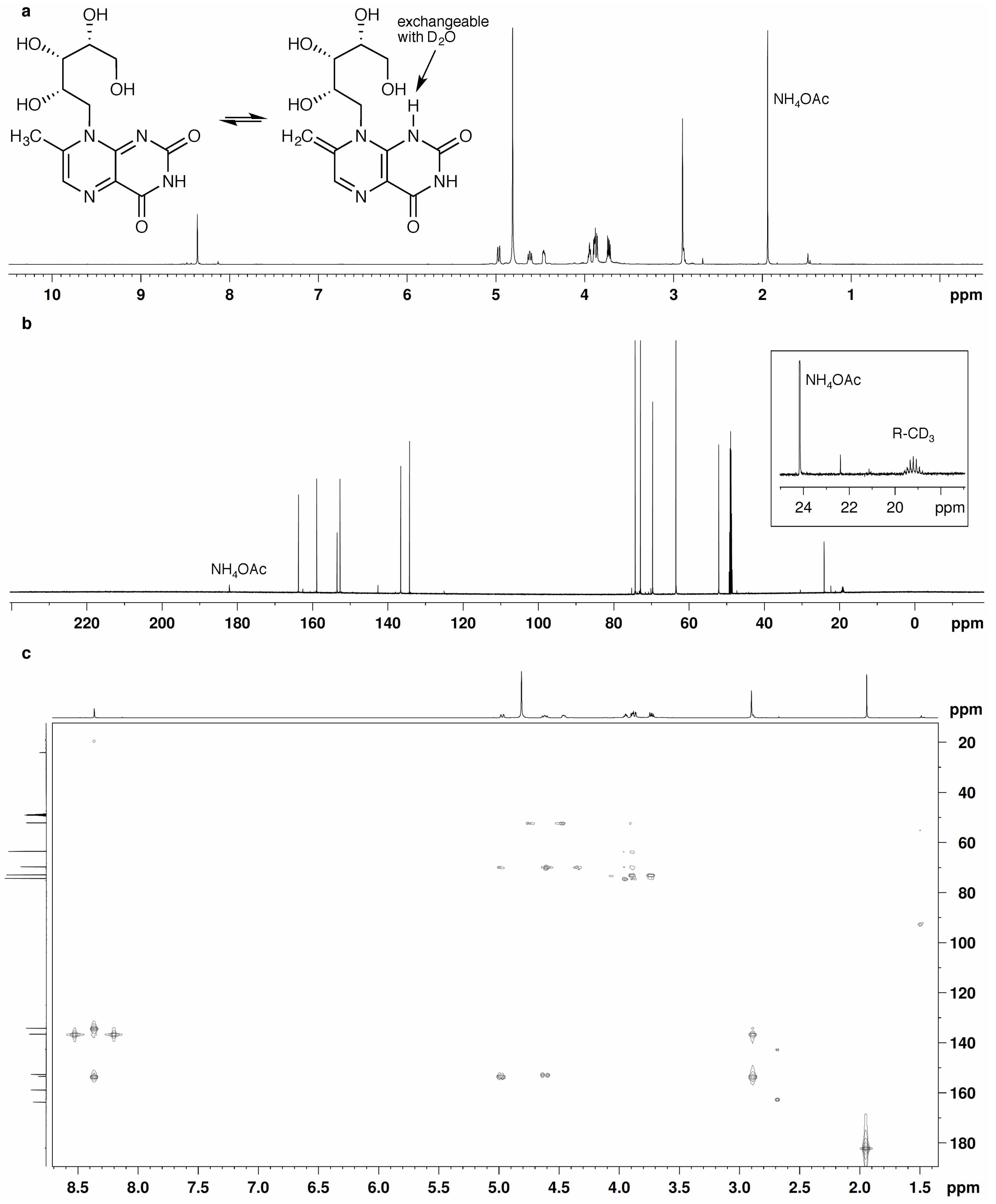
Extended Data Figure 5 | Chromatographic and mass spectrometry properties of MR1 ligands. **a**, Ligand eluted from MR1 complexed with the product of (i) 5-A-RU and methylglyoxal condensation reaction (top); (ii) 5-A-RU and glyoxal condensation reaction (middle); or (iii) 5-A-RU and ^{13}C-glycolaldehyde condensation reaction (bottom). Shown are extracted ion chromatograms (left); *m*/*z* spectrum (centre); and product ions from targeted fragmentation (right). Black diamonds indicate precursor ions. This experiment was performed three times. **b**, Mass spectrometry characterization of 5-OP-RU (top) and 5-OE-RU and ^{13}C-5-OE-RU (bottom).



Extended Data Figure 6 | Mass spectrometry of the 315.09 species. Extracted ion chromatograms of *m*/*z* 315.09 species in MR1 refolded with 5-A-RU and glyoxal (control), or *E. coli* supernatant, or media. Shown are counts on the *y*-axis versus retention time on the *x*-axis. This experiment was performed three times.



Extended Data Figure 7 | NMR characterization of RL. **a–c**, Spectra were recorded as a solution in D_2O–CD_3OD (9:1) with internal solvent peak at 3.31 p.p.m. and 49.0 p.p.m. for ^1H and ^{13}C, respectively. **a**, ^1H NMR (600 MHz); **b**, ^{13}C NMR (150 MHz); **c**, HMBC.



Extended Data Figure 8 | NMR characterization of RL-7-Me. **a**–**c**, Spectra were recorded as a solution in D_2O–CD_3OD (9:1) with internal solvent peak at 3.31 p.p.m. and 49.0 p.p.m. for ^1H and ^{13}C, respectively. **a**, ^1H NMR (600 MHz) and mechanism for deuterium exchange of CH_3 at position 7. Identical exchange was also observed in pure D_2O at slower rate (data not shown). **b**, ^{13}C NMR (150 MHz) showing characteristic heptet from 7-CD_3 after complete deuterium exchange. **c**, HMBC.

Extended Data Table 1 | Data collection and refinement statistics

	MAIT-MR1-5-OP-RU	MAIT-MR1-5-OE-RU	MAIT-K43A-MR1-5-OP-RU
Data collection			
Temperature	100K	100K	100K
Space group	*C*2	*C*2	*C*2
Cell dimensions			
a, b, c (Å)	218.76, 71.11, 144.28	218.11, 70.60, 143.86	215.58, 68.87, 142.98
α, β, γ (°)	90, 104.87, 90	90, 104.63, 90	90, 104.86, 90
Resolution (Å)	33.42-2.50 (2.55-2.50)	75.41-2.10 (2.21-2.10)	50.00-2.20 (2.3-2.20)
R_{pim}*	9.4 (38.8)	6.1 (35.2)	5.9 (36.7)
I/σ_1	7.8 (2.3)	8.1 (2.1)	9.7 (2.3)
Completeness (%)	100 (100)	98.6 (97.1)	97.9 (97.4)
Total N$^{o.}$ observations	307877 (19059)	462978(62837)	509054 (74702)
N$^{o.}$ unique observations	74555 (4584)	122109(17496)	101222 (14608)
Multiplicity	4.1 (4.2)	3.8 (3.6)	5.0 (5.1)
Refinement statistics			
R_{factor} † (%)	16.5	18.4	20.8
R_{free} ‡(%)	21.6	22.2	24.5
No. atoms			
• Protein	12424	12396	12514
• Ligand	45	42	46
• Water	1044	900	488
Ramachandran plot (%)			
• Most favoured	97.4	91.4	91.5
• Allowed region	2.5	8.6	8.5
B-factors (Å^2)			
• Protein	29.8	37.9	37.4
• ligand	14.8	23.8	26
rmsd bonds (Å)	0.010	0.010	0.010
rmsd angles (°)	1.16	1.05	1.08

Values in parentheses refer to the highest-resolution bin.
The R_{factor} was calculated from all data except for 5% that was used for the R_{free} calculation.
* $R_{p.i.m} = \Sigma_{hkl} [1/(N-1)]^{1/2} \Sigma_i \left| I_{hkl,i} - \langle I_{hkl} \rangle \right| / \Sigma_{hkl} \langle I_{hkl} \rangle$
† $R_{factor} = (\Sigma \left| |F_o| - |F_c| \right|) / (\Sigma |F_o|)$; for all data except as indicated by ‡.

LETTER

doi:10.1038/nature13157

Caspase-11 activation requires lysis of pathogen-containing vacuoles by IFN-induced GTPases

Etienne Meunier[1], Mathias S. Dick[1]*, Roland F. Dreier[1]*, Nura Schürmann[1], Daniela Kenzelmann Broz[2], Søren Warming[3], Merone Roose-Girma[3], Dirk Bumann[1], Nobuhiko Kayagaki[3], Kiyoshi Takeda[4], Masahiro Yamamoto[4] & Petr Broz[1]

Lipopolysaccharide from Gram-negative bacteria is sensed in the host cell cytoplasm by a non-canonical inflammasome pathway that ultimately results in caspase-11 activation and cell death[1–3]. In mouse macrophages, activation of this pathway requires the production of type-I interferons[4,5], indicating that interferon-induced genes have a critical role in initiating this pathway. Here we report that a cluster of small interferon-inducible GTPases, the so-called guanylate-binding proteins, is required for the full activity of the non-canonical caspase-11 inflammasome during infections with vacuolar Gram-negative bacteria. We show that guanylate-binding proteins are recruited to intracellular bacterial pathogens and are necessary to induce the lysis of the pathogen-containing vacuole. Lysis of the vacuole releases bacteria into the cytosol, thus allowing the detection of their lipopolysaccharide by a yet unknown lipopolysaccharide sensor. Moreover, recognition of the lysed vacuole by the danger sensor galectin-8 initiates the uptake of bacteria into autophagosomes, which results in a reduction of caspase-11 activation. These results indicate that host-mediated lysis of pathogen-containing vacuoles is an essential immune function and is necessary for efficient recognition of pathogens by inflammasome complexes in the cytosol.

Previous studies have reported that induction of caspase-11-dependent cell death by Gram-negative bacteria requires Trif-dependent production of type-I interferons (type-I-IFNs)[4,5] (Extended Data Fig. 1a). Type-I-IFN production is however not required for pro-caspase-11 induction[4,6,7] and is dispensable for caspase-11 activation by transfected lipopolysaccharide (LPS; Extended Data Fig. 1b)[2]. This indicates that interferon-stimulated genes (ISGs) play a major role in activating caspase-11 in response to intracellular bacteria. To investigate which ISGs were involved in activating caspase-11, we used proteomics-based expression analysis to identify proteins that were highly induced following *Salmonella* infection. Among the most strongly upregulated proteins were interferon-induced GTPases, such as the large 65–67 kDa guanylate-binding proteins (GBPs) and small 47 kDa immunity-related GTPases (IRGs) (data not shown). These proteins function in cell-autonomous immunity, that is, mechanisms that allow host cells to kill pathogens or restrict their replication, and have even been associated with the activation of inflammasomes[8–10].

Mice have 11 GBPs, which are highly homologous and are clustered in two genomic loci on chromosomes 3 and 5, respectively[8,11]. Recently, GBPs on chromosome 3 have been shown to restrict the replication of *Toxoplasma gondii* in peritoneal macrophages and mice[11]. We therefore infected bone-marrow-derived macrophages (BMDMs) from Gbp^{chr3} KO mice, which lack GBP1, 2, 3, 5 and 7 (Extended Data Fig. 2a–e), and wild-type littermates with a number of Gram-negative vacuolar pathogens that trigger caspase-11 activation (data not shown)[1,4,5] and determined the activity of the non-canonical inflammasome pathway at 16 h post-infection (Fig. 1a, b). Macrophages from Gbp^{chr3} KO mice showed a significant reduction of cell death (as measured by lactate dehydrogenase (LDH) release) and IL-1β secretion when infected with wild-type *Salmonella typhimurium*, a type three secretion system (T3SS)-deficient mutant of *S. typhimurium* (ΔSPI-2), *Vibrio cholerae*, *Enterobacter cloacae* or *Citrobacter koseri* (Fig. 1a), and this was independent of LPS or polyinosinic:polycytidylic acid (poly(I:C)) priming (Extended Data Fig. 2f, g). Gbp^{chr3}-deficiency also reduced secretion of caspase-1 p20 subunit, caspase-11 and mature IL-1β, IL-18 and IL-1α (Fig. 1b). Because interferons induce GBP expression (Extended Data Fig. 2b, c)[8], we investigated whether IFN-γ treatment would accelerate LDH release in response to *Salmonella* infection. IFN-γ-treated wild-type BMDMs released LDH as soon as 4 h after infection, whereas Gbp^{chr3} KO BMDMs failed to release LDH at early time points even after IFN-γ priming (Fig. 1c), indicating that GBP induction was required for activity of the non-canonical inflammasome pathway.

We next explored whether GBPs play a role in the activation of canonical inflammasomes. LPS-primed wild-type and Gbp^{chr3}-deficient macrophages released comparable levels of LDH and mature IL-1β when infected with logarithmic phase *S. typhimurium*, which exclusively engage the NLRC4 inflammasome via the SPI-1 T3SS (Fig. 1d)[12]. Similarly, Gbp^{chr3}-deficiency did not affect AIM2 inflammasome activation upon poly (deoxyadenylic-deoxythymidylic) acid (poly(dA:dT)) transfection (Fig. 1d). Although GBP5 had been previously linked to NLRP3 activation[9], we did not observe a defect in NLRP3 activation in Gbp^{chr3} KOs (Fig. 1d), possibly owing to different modes of pre-stimulation. These data indicate that GBPs are dispensable for canonical inflammasome activity, but are required for the activation of the non-canonical inflammasome pathway.

To investigate whether GBPs directly mediated the detection of intracellular LPS, we engaged the non-canonical inflammasome by transfecting macrophages with different types of ultra-pure LPS (Fig. 1e). Cytoplasmic LPS triggered LDH release and IL-1β secretion to a similar extent in both wild-type and Gbp^{chr3}-deficient BMDMs, indicating that GBPs were required upstream of LPS sensing and only during bacterial infection. We next investigated if GBPs were required for immune detection of vacuolar or cytosolic bacteria by infecting BMDMs with Δ*sifA S. typhimurium* and *Burkholderia thailandensis*, which rapidly enter the cytosol and activate caspase-11 (ref. 13). Unprimed Gbp^{chr3} KO and wild-type BMDMs responded comparably to these bacteria (Extended Data Fig. 3a–c). Because GBPs might affect this response when pre-induced, we also infected IFN-γ-primed BMDMs with Δ*sifA S. typhimurium* (Extended Data Fig. 3d). IFN-γ-priming indeed resulted in a small difference between wild-type and Gbp^{chr3} KO BMDMs after infection with Δ*sifA Salmonella*, yet not to the extent seen with wild-type *Salmonella* (Fig. 1c), indicating that GBPs mainly participate in the activation of the non-canonical inflammasome by vacuolar bacteria.

Finally, to investigate which GBP controls caspase-11 activation, all 11 murine *Gbp*s were individually knocked down in BMDMs and the cells were infected with flagellin-deficient *Salmonella*, which activate the non-canonical inflammasome but not NLRC4 (Extended Data Fig. 4a and Supplementary Information)[4]. Only knockdown of *Gbp2* resulted in reduced LDH release and IL-1β secretion (Extended Data Fig. 4b–d).

[1]Focal Area Infection Biology, Biozentrum, University of Basel, CH-4056 Basel, Switzerland. [2]Department Biomedicine, University of Basel, CH-4056 Basel, Switzerland. [3]Genentech Inc., South San Francisco, California 94080, USA. [4]Department of Microbiology and Immunology, Osaka University, Yamadaoka, Suita, Osaka 565-0871, Japan.
*These authors contributed equally to this work.

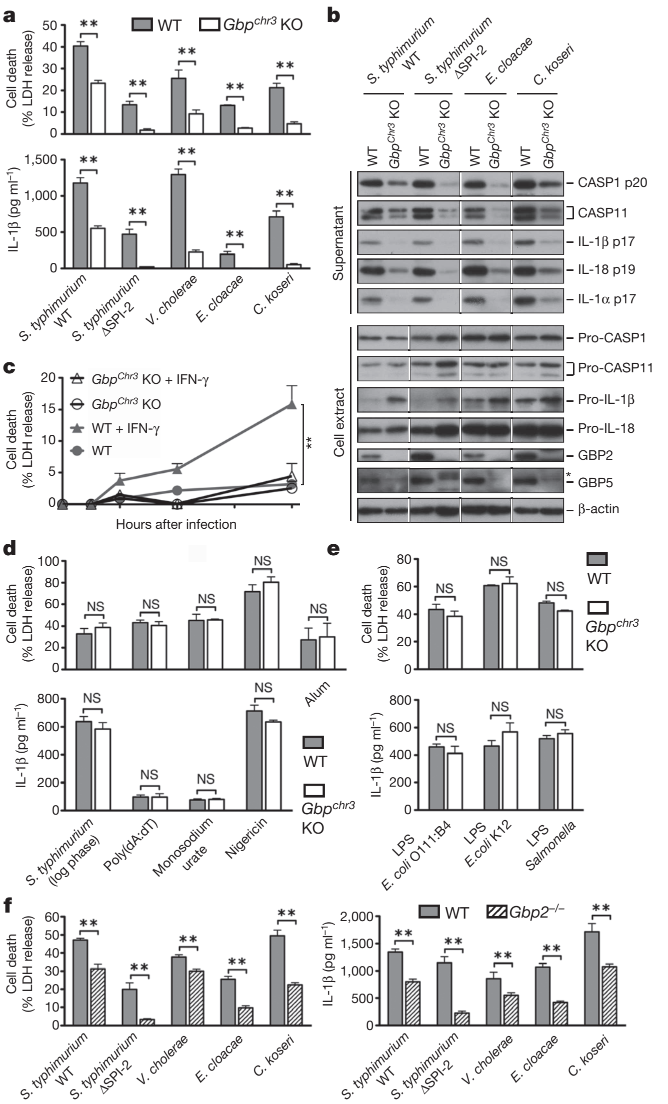


Figure 1 | Caspase-11 activation by intracellular bacterial pathogens requires GBPs. **a**, **b**, LDH release, IL-1β secretion (**a**) and immunoblots for caspase-1, caspase-11, IL-1β, IL-18 and IL-1α (**b**) from unprimed BMDMs infected for 16 h with the indicated bacteria (grown to stationary phase). **c**, Time course measuring LDH release from unprimed or IFN-γ-primed BMDMs infected with *S. typhimurium*. **d**, **e**, LDH release and IL-1β secretion from primed BMDMs infected with SPI-1-expressing logarithmic phase *S. typhimurium*, treated with monosodium urate, alum and nigericin or transfected with poly(dA:dT) and LPS. **f**, LDH release and IL-1β secretion from unprimed wild-type and $Gbp2^{-/-}$ BMDMs infected for 16 h with the indicated bacteria (grown to stationary phase). Graphs show mean and s.d. of quadruplicate wells and data are representative of two (**b**) and three (**a**, **c**–**f**) independent experiments. *Crossreactive band; $^{**}P < 0.01$; NS, not significant (two-tailed *t*-test).

To validate these data we obtained BMDMs from $Gbp2^{-/-}$ mice and wild-type littermates[14] and infected them with vacuolar Gram-negative bacteria. As expected, we observed reduced levels of cell death, cytokine secretion and caspase release in $Gbp2^{-/-}$ BMDMs, indicating attenuated activation of the non-canonical inflammasome (Fig. 1f and Extended Data Fig. 4e), whereas direct LPS sensing or the activation of canonical inflammasomes was not affected (Extended Data Fig. 4f, g). In contrast, *Gbp5*-deficiency did not have any effect on canonical and non-canonical inflammasome activation (Extended Data Fig. 5). Nevertheless, *Gbp2*-deficiency did not reduce caspase-11 activation as markedly as Gbp^{chr3}-deficiency, indicating that whereas caspase-11 activation mainly requires GBP2, other GBPs might also be partially involved.

Reduced numbers of intracellular bacteria could account for low levels of caspase-11 activation in Gbp^{chr3}- and *Gbp2*-deficient macrophages. However, a comparison of wild-type and Gbp^{chr3} KO BMDMs showed that Gbp^{chr3}-deficiency resulted in significantly higher numbers of total and live *Salmonella* per cell (Fig. 2a), consistent with higher colony forming units numbers in Gbp^{chr3} KO BMDMs (Extended Data Fig. 6). In addition, fluorescence-activated cell sorting (FACS)-based analysis of dead (mCherry-negative, FITC$^+$) and live (mCherry-positive, FITC$^+$) *Salmonella* at 16 h post-infection found significantly fewer dead bacteria (~20%) in GBP^{chr3} KO and $Gbp2^{-/-}$ BMDMs when compared to wild-type BMDMs (>30%) (Fig. 2b). Importantly, bacterial killing in $Casp11^{-/-}$ BMDMs was comparable to wild-type BMDMs, indicating that the control of bacterial replication was directly linked to GBP function and not to the activation of the non-canonical inflammasome (Fig. 2b). In conclusion, we show that GBPs control bacterial replication on a cell-autonomous level, which is consistent with a previous report that GBP1 partially restricts *Mycobacterium bovis* and *Listeria monocytogenes* replication[10].

Restricting bacterial replication has been proposed to require the association of GBPs with pathogen-containing vacuoles and the recruitment of antimicrobial factors[8]. We therefore investigated whether GBPs targeted intracellular Gram-negative bacteria. Indeed, GBP2 could be detected on intracellular bacteria within hours after infection (Fig. 2c). Very little GBP-positive bacteria were detected in $Stat1^{-/-}$ BMDMs, which do not respond to type-I- and type-II-IFNs and largely failed to induce GBP expression (data not shown). Remarkably, GBP-positive *Salmonella* seemed to have lost mCherry expression (Fig. 2c), indicating that these bacteria were dead. To determine whether GBPs are recruited to dead bacteria we infected BMDMs with *Salmonella* killed by heat, paraformaldehyde or 70% ethanol treatment, yet only live *Salmonella* acquired GBP staining and activated the inflammasome (Fig. 2d). To examine this mechanism *in vivo*, we immunostained spleen tissue sections of mice infected with *Salmonella* for GBPs. Indeed, GBPs could also be found associated with approximately 20% of bacteria *in vivo*, and a significantly higher proportion of these bacteria were dead, based on the loss of mCherry expression (Fig. 2e–g). Furthermore, treatment with IFN-γ-neutralizing antibodies reduced the percentage of GBP-positive bacteria (Fig. 2f), consistent with reports that IFN-γ controls *Salmonella* replication *in vivo*[15,16]. Taken together, these results indicated that GBPs either kill bacteria directly or control an antimicrobial effector pathway, and raised the interesting possibility that GBP-mediated killing of bacteria might result in the release of LPS and caspase-11 activation[2,3].

To identify the antimicrobial effector pathway that is controlled by GBPs we first examined the role of free radicals[8]. Although GBP7 was reported to be required for reactive oxygen species (ROS) production and to interact with the phagosome oxidase complex[10], we did not find any role for ROS or NO production in caspase-11 activation (Extended Data Fig. 7). Furthermore, GBPs were also proposed to recruit components of the autophagy machinery to pathogen-containing vacuoles (PCVs), possibly resulting in bacterial killing within autophagosomes[8,10]. Indeed, many GBP-positive *S. typhimurium*, *E. cloacae* and *C. koseri* co-stained for the commonly used autophagy marker LC3 (Fig. 3a and Extended Data Fig. 8a). Recruitment of LC3 to intracellular *Salmonella* was partially GBP-dependent, because we found significantly lower numbers of LC3-positive *Salmonella* in Gbp^{chr3} KO compared to wild-type macrophages (Fig. 3b, c). Therefore, we speculated that autophagy-mediated killing might result in the release of LPS from bacteria and caspase-11 activation. Unexpectedly, however, pharmacological inhibition of autophagy with 3-methyladenine (3-MA) resulted in significantly higher levels of LDH release, IL-1β secretion and caspase-1/caspase-11 activation in macrophages infected with *S. typhimurium*, *E. cloacae* or *C. koseri* (Fig. 3d, e), indicating increased activation of the non-canonical inflammasome. Consistently, cell death was still caspase-11-dependent because $Casp11^{-/-}$ BMDMs did not release LDH when treated with 3-MA and infected with Gram-negative bacteria (Fig. 3f). Direct activation of caspase-11 by LPS transfection was independent of

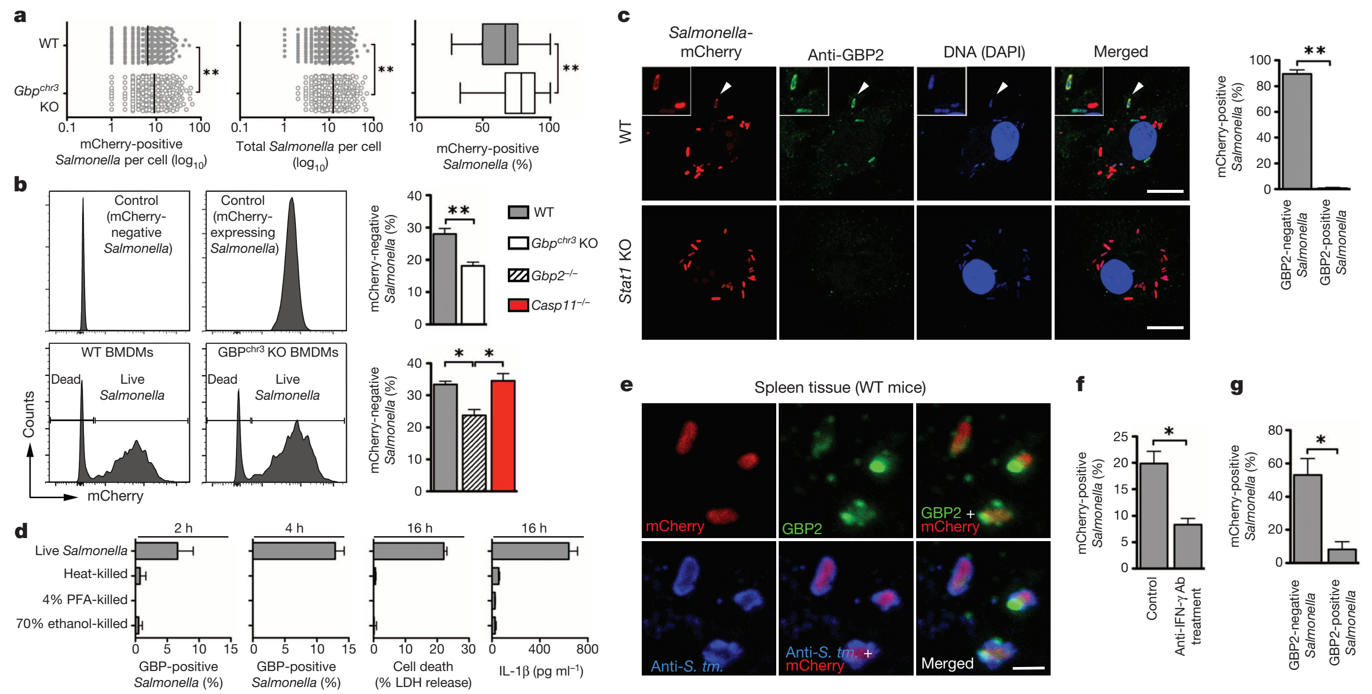


Figure 2 | GBPs control bacterial replication. **a**, **b**, Quantification of live (mCherry-positive) and dead (mCherry-negative) *S. typhimurium* per cell by immunofluorescence (**a**) or as percent of total by flow-cytometry (**b**) in unprimed BMDMs at 16 h post-infection. **c**, Immunostaining for GBP2 and quantification of live and dead *Salmonella* at 4 h post-infection. Arrowheads, bacteria shown in insets. **d**, Quantification of GBP-positive bacteria, LDH release and IL-1β secretion at indicated time points from BMDMs infected with *Salmonella*, live or killed by different means. **e**, Immunohistochemistry for GBP2 and *Salmonella* on spleen tissue from *Salmonella* (mCherry-positive)-infected mice (representative of $n = 3$ per group). *S. tm.*, *S. typhimurium*. **f**, **g**, Quantification of GBP-positive *Salmonella* in anti-IFN-γ-treated or control animals (**f**) and live and dead bacteria among GBP2-negative/-positive *Salmonella* (**g**) ($n = 3$ per group). Scale bars, 10 μm (**c**), 1 μm (**e**). Graphs show mean and 5–95 percentile (box plots) or s.d. of technical triplicates, and data are representative of three independent experiments. $^*P < 0.05$, $^{**}P < 0.01$ (two-tailed *t*-test).

autophagy (Fig. 3g), indicating that autophagy only counteracts non-canonical inflammasome activation during bacterial infections. To further confirm our data, we infected *Atg5*$^{-/-}$ BMDMs with *S. typhimurium* and we also observed significantly higher levels of non-canonical inflammasome activation compared to wild-type BMDMs (Fig. 3h, i). Taken together, these results indicated that, although GBPs promoted the uptake of bacteria into autophagosomes, autophagy actually counteracted caspase-11 activation. Thus, GBP-dependent LPS detection occurs before bacteria are targeted to autophagosomes.

A possible explanation could be that autophagy sequesters bacteria that had escaped from the vacuole, and thus prevents further LPS release into the cytosol. Recently, the cytosolic danger receptor galectin-8 was reported to function as a marker for lysed vacuoles. Galectin-8 binds β-galactosides, which are normally found on the inner leaflet of the vacuolar membrane and get exposed to the cytosol upon vacuolar lysis[17]. Indeed, quantification of galectin-8-positive *Salmonella* showed that significantly fewer bacteria were targeted by galectin-8 in *Gbp*^{chr3} KO BMDMs than in wild-type macrophages (Fig. 4a). Because galectin-8 colocalized with GBP- and LC3-positive *Salmonella* (Fig. 4b, c), we speculated that GBPs promote LC3 recruitment through galectin-8. Consistently, we found lower levels of galectin-8-positive *Salmonella* among LC3-positive *Salmonella* in *Gbp*^{chr3} KO compared to wild-type BMDMs (Fig. 4d). Galectin-8 interacts with the autophagy adaptor protein NDP52, which in humans contains binding sites for galectin-8, ubiquitin and LC3[18]. In line with a role for NDP52 in linking galectin-8 to LC3, murine NDP52 colocalized with galectin-8 on intracellular *Salmonella* (Extended Data Fig. 8b). Targeting of *Salmonella* to autophagosomes might also involve other autophagy cargo adaptors, because p62 was associated with the majority of LC3-positive bacteria, yet this was independent of GBPs (Extended Data Fig. 8c, d). Altogether, these results suggested that GBPs might promote the lysis of vacuoles or help to recruit galectin-8 to lysed vacuoles.

To confirm a direct role of GBPs in vacuolar lysis, we adapted a phagosome integrity assay based on differential permeabilization with digitonin (Extended Data Fig. 9). Comparing wild-type and *GBP*^{chr3} KO BMDMs, we found significantly lower numbers of cytosolic (FITC$^+$) *S. typhimurium* in *Gbp*^{chr3}-deficient cells (Fig. 4e, f). Similarly, *Gbp2*$^{-/-}$ BMDMs also harboured fewer cytosolic *S. typhimurium* compared to BMDMs from wild-type littermates (Fig. 4g). In contrast, we did not find a defect in cytosolic localization between wild-type and *Gbp*^{chr3} KO BMDMs infected with the specialized cytosolic pathogen *Shigella flexneri*, which uses its T3SS to destabilize the phagosome and escape into the cytoplasm (Fig. 4h)[19]. Although we cannot exclude that GBPs might also be involved in the recruitment or assembly of the non-canonical inflammasome, these results indicate that GBPs, in particular GBP2, directly promote the destruction of vacuoles.

In conclusion, our data demonstrate that host-induced destruction of PCVs or phagosomes is an essential immune function and assures recognition of vacuolar bacteria by cytosolic innate immune sensors (Extended Data Fig. 10). Additional studies are required to determine how GBPs distinguish 'self' and 'non-self' membranes and by which mechanism phagosomes are lysed. In mice, this might involve the IRGM proteins that can act as GDI (guanine nucleotide dissociation inhibitor) and inhibit IRG and GBP activity. Absence of IRGMs results in mislocalization of both IRGs and GBPs and even in degradation of lipid droplets[20–22], supporting a model in which IRGM proteins would protect 'self'-vacuoles from being targeted by host IRGs and GBPs[23]. Because both commensals and pathogens activate caspase-11 (ref. 1), it can be assumed that GBPs are not specific towards pathogens but are a general innate immune response against bacteria trapped in the phagosomes of macrophages. Finally, given the important role of LPS-induced caspase-11 activation in septic shock[1–3], pharmaceutical targeting of the above-described pathways might be used to modulate inflammation during bacterial sepsis.

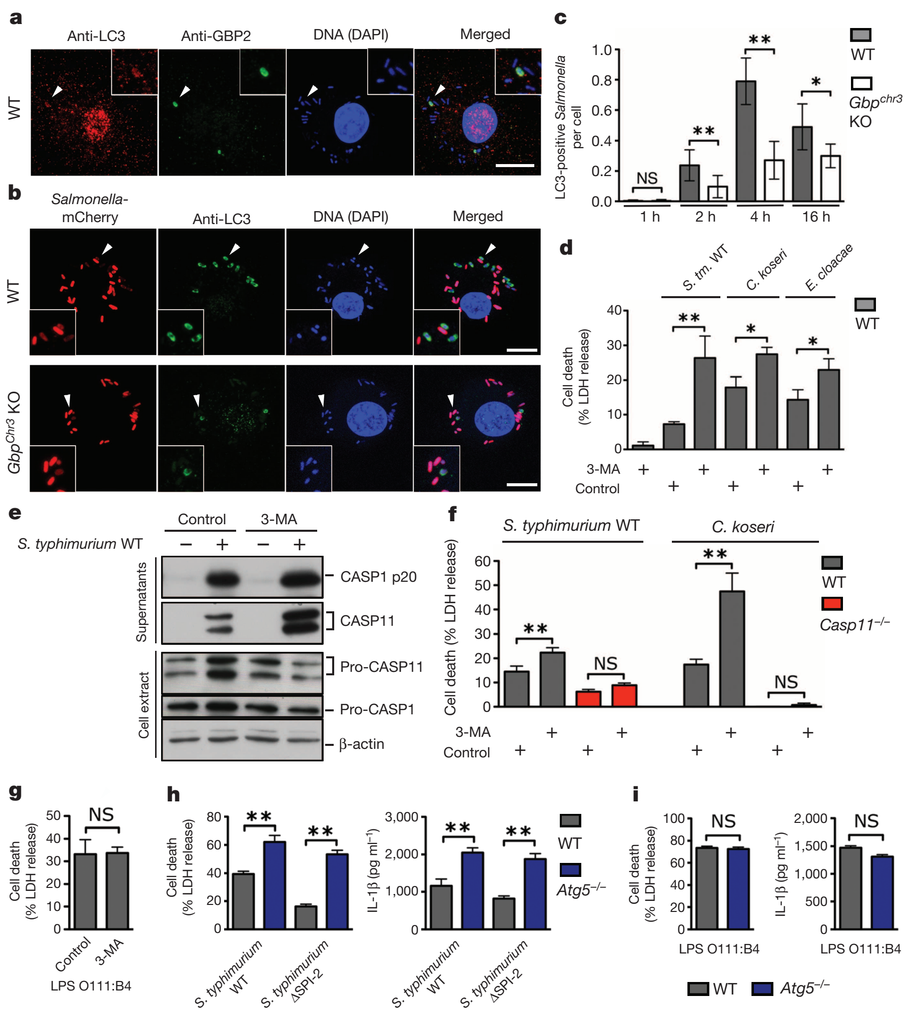


Figure 3 | Autophagy reduces caspase-11 activation. **a**, **b**, Unprimed BMDMs infected with *S. typhimurium* for 4 h and immunostained for LC3 and GBP2. Arrowheads, bacteria shown in insets. Scale bars, 10 µm. **c**, Quantification of results from **b**. **d**–**g**, LDH release and immunoblots for caspase-1 and caspase-11 from BMDMs infected for 16 h or transfected with LPS in presence or absence of 3-methyladenine (3-MA). **h**, **i**, LDH release and IL-1β secretion from BMDMs infected for 16 h or transfected with LPS. Graphs show mean and s.d. of quadruplicate wells and data are representative of two (**e**, **i**) and three (**a**–**d**, **f**–**h**) independent experiments. *$P < 0.05$, **$P < 0.01$; NS, not significant (two-tailed *t*-test).

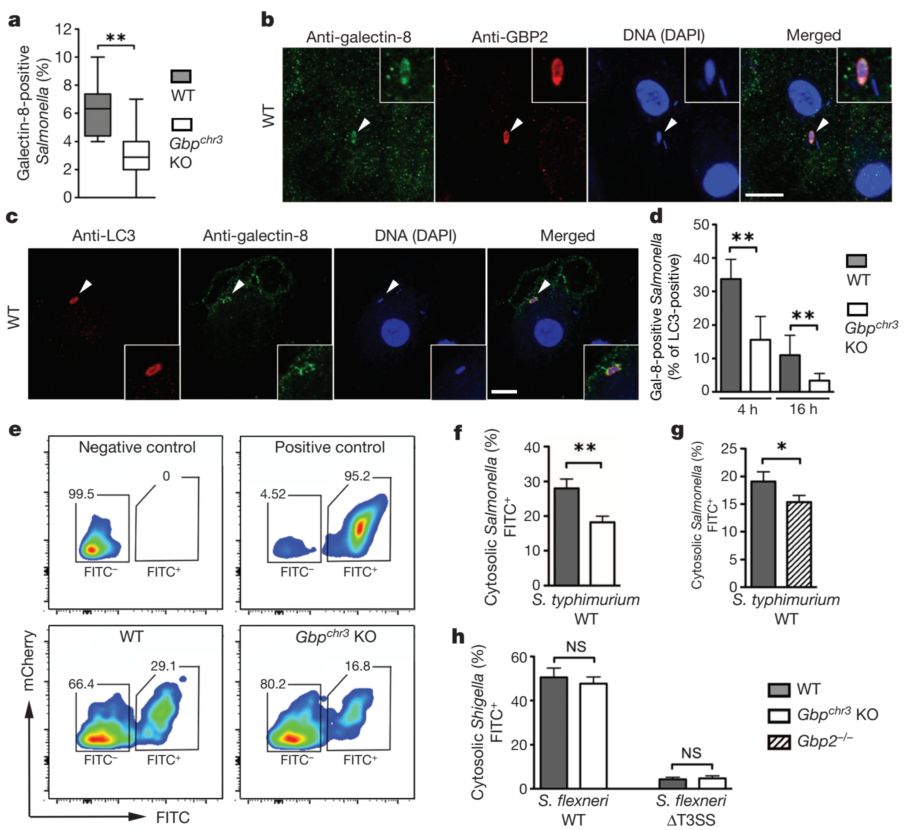


Figure 4 | GBP-mediated lysis of the PCV releases *Salmonella* into the cytosol. **a**, Quantification of galectin-8-positive *Salmonella* in unprimed BMDMs at 4 h post-infection. **b**, **c**, Unprimed BMDMs infected with *S. typhimurium* for 4 h and immunostained for galectin-8, GBP2 and LC3. Arrowheads, bacteria shown in insets. Scale bars, 10 µm. **d**, Quantification of galectin-8/LC3-double-positive *Salmonella* at indicated time points post-infection. **e**–**h**, Quantification of cytosolic and vacuolar bacteria by flow cytometry in BMDMs infected with mCherry-positive *S. typhimurium* (**e**–**g**) or *S. flexneri* (**h**, wild-type or ΔT3SS) for 4 h. Graphs show mean and s.d. or 5–95 percentile (Box plots) of technical triplicates. Data are representative of 2 (**g**, **h**), 3 (**a**–**d**) and 4 (**e**, **f**) independent experiments. *$P < 0.05$, **$P < 0.01$ (two-tailed *t*-test).

METHODS SUMMARY

BMDMs were cultured and seeded for infections as described previously[4]. Priming was done overnight with PAM3CSK4 (1 μg ml^{-1}), LPS O111:B4 (0.1 μg ml^{-1}), murine IFN-β or murine IFN-γ (1 unit per μl). *S. typhimurium*, *S. flexneri*, *V. cholerae*, *E. cloacae*, *C. koseri* and *B. thailandensis* were grown overnight in LB or TSB medium at 37 °C with aeration. Bacteria were diluted in fresh pre-warmed macrophage medium and added to the macrophages at a multiplicity of infection (m.o.i.) of 100:1 for measurements of caspase-11 and caspase-1 activity or 10:1 for all other assays. For assaying NLRC4 activation, *Salmonella* were subcultured for 4 h to induce SPI-1 T3SS expression before infection (m.o.i. 20:1). *S. flexneri* were subcultured for 3 h to induce T3SS expression before infection (m.o.i. 30:1). When required, apocynin, L-NG-nitroarginine methyl ester (L-NAME), 3-methyladenine or vehicle controls were added 30 min before infection. Plates were centrifuged for 15 min at 500*g* to synchronize the infection and placed at 37 °C for 1 h. Next, 100 μg ml^{-1} gentamycin was added to kill extracellular bacteria. After 1 h incubation, the cells were washed once with DMEM and given fresh macrophage medium containing 10 μg ml^{-1} gentamicin for the remainder of the infection. Transfection with poly(dA:dT) or MSU, alum or nigericin treatment was done as described previously[2] or as indicated. All animal experiments were approved and performed according to local guidelines. Female BALB/c mice (10–14 weeks old) were infected intravenously with *Salmonella* (1,000 c.f.u.) and euthanized 4–5 days later. For antibody injections, mice received on day 3 two intraperitoneal injections of 200 μl PBS containing 0.2 mg anti-IFN-γ monoclonal or 0.2 mg rat IgG1, κ isotype control antibody.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 31 October 2013; accepted 14 February 2014.
Published online 16 April 2014.

1. Kayagaki, N. *et al.* Non-canonical inflammasome activation targets caspase-11. *Nature* **479,** 117–121 (2011).
2. Kayagaki, N. *et al.* Noncanonical inflammasome activation by intracellular LPS independent of TLR4. *Science* **341,** 1246–1249 (2013).
3. Hagar, J. A., Powell, D. A., Aachoui, Y., Ernst, R. K. & Miao, E. A. Cytoplasmic LPS activates caspase-11: implications in TLR4-independent endotoxic shock. *Science* **341,** 1250–1253 (2013).
4. Broz, P. *et al.* Caspase-11 increases susceptibility to *Salmonella* infection in the absence of caspase-1. *Nature* **490,** 288–291 (2012).
5. Rathinam, V. A. *et al.* TRIF licenses caspase-11-dependent NLRP3 inflammasome activation by Gram-negative bacteria. *Cell* **150,** 606–619 (2012).
6. Case, C. L. *et al.* Caspase-11 stimulates rapid flagellin-independent pyroptosis in response to *Legionella pneumophila*. *Proc. Natl Acad. Sci. USA* **110,** 1851–1856 (2013).
7. Casson, C. N. *et al.* Caspase-11 activation in response to bacterial secretion systems that access the host cytosol. *PLoS Pathog.* **9,** e1003400 (2013).
8. MacMicking, J. D. Interferon-inducible effector mechanisms in cell-autonomous immunity. *Nature Rev. Immunol.* **12,** 367–382 (2012).
9. Shenoy, A. R. *et al.* GBP5 promotes NLRP3 inflammasome assembly and immunity in mammals. *Science* **336,** 481–485 (2012).
10. Kim, B. H. *et al.* A family of IFN-gamma-inducible 65-kD GTPases protects against bacterial infection. *Science* **332,** 717–721 (2011).
11. Yamamoto, M. *et al.* A cluster of interferon-gamma-inducible p65 GTPases plays a critical role in host defense against *Toxoplasma gondii*. *Immunity* **37,** 302–313 (2012).
12. Broz, P. *et al.* Redundant roles for inflammasome receptors NLRP3 and NLRC4 in host defense against *Salmonella*. *J. Exp. Med.* **207,** 1745–1755 (2010).
13. Aachoui, Y. *et al.* Caspase-11 protects against bacteria that escape the vacuole. *Science* **339,** 975–978 (2013).
14. Degrandi, D. *et al.* Murine guanylate binding protein 2 (mGBP2) controls *Toxoplasma gondii* replication. *Proc. Natl Acad. Sci. USA* **110,** 294–299 (2013).
15. VanCott, J. L. *et al.* Regulation of host immune responses by modification of *Salmonella* virulence genes. *Nature Med.* **4,** 1247–1252 (1998).
16. Burton, N. A. *et al.* Disparate impact of oxidative host defenses determines the fate of *Salmonella* during systemic infection in mice. *Cell Host Microbe* **15,** 72–83 (2014).
17. Thurston, T. L., Wandel, M. P., von Muhlinen, N., Foeglein, A. & Randow, F. Galectin 8 targets damaged vesicles for autophagy to defend cells against bacterial invasion. *Nature* **482,** 414–418 (2012).
18. Deretic, V., Saitoh, T. & Akira, S. Autophagy in infection, inflammation and immunity. *Nature Rev. Immunol.* **13,** 722–737 (2013).
19. Paetzold, S., Lourido, S., Raupach, B. & Zychlinsky, A. *Shigella flexneri* phagosomal escape is independent of invasion. *Infect. Immun.* **75,** 4826–4830 (2007).
20. Hunn, J. P. *et al.* Regulatory interactions between IRG resistance GTPases in the cellular response to *Toxoplasma gondii*. *EMBO J.* **27,** 2495–2509 (2008).
21. Traver, M. K. *et al.* Immunity-related GTPase M (IRGM) proteins influence the localization of guanylate-binding protein 2 (GBP2) by modulating macroautophagy. *J. Biol. Chem.* **286,** 30471–30480 (2011).
22. Haldar, A. K. *et al.* IRG and GBP host resistance factors target aberrant, "non-self" vacuoles characterized by the missing of "self" IRGM proteins. *PLoS Pathog.* **9,** e1003414 (2013).
23. Coers, J. Self and non-self discrimination of intracellular membranes by the innate immune system. *PLoS Pathog.* **9,** e1003538 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank N. Mizushima and S. Virgin for *Atg5*-deficient BMDMs, K. Pfeffer for *Gbp2*-deficient BMDMs, J. Frey for *B. thailandensis*, the Biozentrum Proteomics and Imaging Core Facilities for technical assistance, K. Anderson, T. Soukup, R. Schwingendorf, J. C. Cox, V. M. Dixit for reagents and N. Personnic for discussions. This work was supported by an SNSF Professorship PP00P3_139120/1, University of Basel project grant ID2153162 to P.B. and a Marie Heim-Voegtlin Fellowship 145516 to D.K.B.

Author Contributions E.M. and P.B. designed the study and wrote the manuscript. E.M., R.F.D., M.S.D., N.S. and P.B. performed the experiments and analysed data; D.K.B., D.B., S.W., M.R.-G., N.K., M.Y. and K.T. contributed reagents.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.B. (petr.broz@unibas.ch).

METHODS

Bacterial strains and plasmids. *Salmonella enterica* serovar Typhimurium (*S. typhimurium*) SL1344 and congenic mutants were published before[12]. Other bacterial strains used were *Shigella flexneri*, *Vibrio cholerae*, *Enterobacter cloacae*, *Citrobacter koseri* and *Burkholderia thailandensis* ATCC700388.

Mice. Gbp^{chr3} KO, $Gbp2^{-/-}$, $Atg5^{fl/fl}$-Lyz-Cre, $Cybb^{-/-}$ ($gp91^{phox}$), $Casp1^{-/-}/Casp11^{-/-}$ (a.k.a caspase-1 knockout), $Casp11^{-/-}$ and $Casp1^{-/-}$ ($Casp1^{-/-}/Casp11^{tg}$) mice have been previously described[1,11,14,24,25]. Mice were bred in the animal facilities of the University of Basel, Genentech Inc., Heinrich-Heine-University Duesseldorf or the University of Osaka. Generation of mice with *Gbp5* KO alleles by zinc finger nuclease (ZFN) technology: A ZFN pair was obtained from Sigma-Aldrich (SAGE Labs). The ZFN pair recognizes a sequence in mouse *Gbp5* exon 2 (cut site is underlined): 5′-TGCCATCACACAGCCAGTGGTGGTGGTAGCCATTGTGGGT-3′. ZFN mRNA and a donor plasmid harbouring a 10-bp deletion in *Gbp5* exon 2 was co-microinjected into C57BL/6N zygotes using established procedures. One male founder carrying the 10-bp deletion was obtained by homologous recombination (10-bp deletion is underlined): 5′-TGCCATCACACAGCCAGTGGTGGTGGTAGCCATTGTGGGT-3. This founder was bred with C57BL/6N females to generate heterozygous progeny for subsequent intercrossing. Two founders (a male and a female) carrying identical 1-bp deletions were obtained by non-homologous end-joining (deleted bp is underlined): 5′-TGCCATCACACAGCCAGTGGTGGTGGTAGCCATTGTGGGT-3. These two founders were intercrossed to directly generate homozygous progeny. Both the 10-bp (designated KO line 1) and 1 bp (designated KO line 2) deletions lead to frameshifts and premature stop codons in *Gbp5* exon 2.

Animal infection. All animal experiments were approved (license 2239, Kantonales Veterinäramt Basel-Stadt) and performed according to local guidelines (Tierschutz-Verordnung, Basel-Stadt) and the Swiss animal protection law (Tierschutz-Gesetz). Female BALB/c mice (10–14 weeks old) were infected intravenously with mCherry-positive *Salmonella* (1,000 c.f.u.) and euthanized 4–5 days later. For antibody injections, mice ($n = 3$ per group) received on day 3 two intraperitoneal injections of 200 μl PBS containing 0.2 mg anti-IFN-γ monoclonal antibody (Clone XMG1.2, BioLegend) or 0.2 mg rat IgG1, κ isotype control antibody (clone RTK2071, BioLegend). No randomization or blinding was performed.

Cell culture and infections. BMDMs were differentiated in DMEM (Invitrogen) with 10% v/v FCS (Thermo Fisher Scientific), 10% MCSF (L929 cell supernatant), 10 mM HEPES (Invitrogen), and nonessential amino acids (Invitrogen). 1 day before infection, macrophages were seeded into 6-, 24-, or 96-well plates at a density of 1.25×10^6, 2.5×10^5, or 5×10^4 per well. If required macrophages were prestimulated with PAM3CSK4, LPS O111:B4 (InvivoGen), mIFN-β or mIFN-γ (eBioscience). For infections with *S. typhimurium*, *V. cholerae*, *E. cloacae*, *C. koseri* and *B. thailandensis*, bacteria were grown overnight in LB or TSB at 37 °C with aeration. The bacteria were diluted in fresh pre-warmed macrophage medium and added to the macrophages at an multiplicity of infection (m.o.i.) of 100:1 for measurements of caspase-11 and caspase-1 activity or 10:1 for all other assays. For assaying *Salmonella*-induced NLRC4 activation, *Salmonella* were subcultured for 4 h before infection to induce SPI-1 T3SS and flagellin expression. *S. flexneri* were cultured overnight in TSB medium and subcultured for 3 h before infection to induce T3SS expression. IFN-γ-primed BMDMs (to induce GBP expression) were infected with m.o.i. of 30:1 with *S. flexneri* for FACS analysis. When required, chemical reagents, Apocynin (Sigma Aldrich, 100μM), L-NG-nitroarginine methyl ester (L-NAME; Sigma Aldrich, 100 μM) and 3-methyladenine (Sigma Aldrich, 5 mM) were added 30 min before infection. The plates were centrifuged for 15 min at 500 *g* to ensure comparable adhesion of the bacteria to the cells and placed at 37 °C for 60 min. Next, 100 μg ml^{-1} gentamycin (Invitrogen) was added to kill extracellular bacteria. After a 60-min incubation, the cells were washed once with DMEM and given fresh macrophage medium containing 10 μg ml^{-1} gentamicin for the remainder of the infection. For infections with killed bacteria, *Salmonella* were grown as above. Shortly before the infection, bacteria were left untreated or incubated for 30 min at 95 °C, in 4% paraformaldehyde or in 70% ethanol. Following the treatment, bacteria were washed with PBS and prepared for infections as outlined above. The effectiveness of the killing procedures was verified by plating serial dilutions. Transfection with poly(dA:dT) or treatment with MSU, alum or nigericin was done as described previously[2] or as indicated.

siRNA knockdown. Gene knockdown was done using GenMute (SignaGen) and siRNA pools (siGenome, Dharmacon). Briefly, wild-type BMDMs were seeded into 24-, or 96-well plates at a density of 1.5×10^5 or 3×10^4 per well. siRNA complexes were prepared at 25 nM siRNA in 1× GenMute Buffer according to the manufacturer's instructions for forward knockdowns. siRNA complexes were mixed with BMDM medium and added onto the cells. BMDMs were infected with *S. typhimurium* at an m.o.i. of 100:1 after 56 h of knockdown and analysed for inflammasome activation as outlined below. siRNA pools included: *Casp11* (that is, *Casp4*) (M-042432-01), *Gbp1* (M-040198-01), *Gbp2* (M-040199-00), *Gbp3* (M-063076-01), *Gbp4* (M-047506-01), *Gbp5* (M-054703-01), *Gbp6* (M-041286-01), *Gbp7* (M-061204-01), *Gbp8* (M-059726-01), *Gbp9* (M-052281-01), *Gbp10* (M-073912-00), *Gbp11* (M-079932-00) and NT (non-targeting) pool 2 (D-001206-14). See Supplementary information for sequences.

LPS transfection. Macrophages were seeded as described above. Cells were prestimulated with 10 ug ml^{-1} of PAM3CSK4 for 4 h in Opti-MEM and transfected for 16 h with ultrapure LPS *E. coli* O111:B4, ultrapure LPS *E. coli* K12 or ultrapure LPS *Salmonella minnesota* (InvivoGen) in complex with FuGeneHD (Promega) as described previously[2].

Cytokine and LDH release measurement. IL-1β and tumour necrosis factor (TNF)-α was measured by ELISA (eBioscience). LDH was measured using LDH Cytotoxicity Detection Kit (Clontech). To normalize for spontaneous lysis, the percentage of LDH release was calculated as follows: (LDH infected − LDH uninfected)/(LDH total lysis − LDH uninfected)*100.

Western blotting. Western blotting was done as described before[4]. Antibodies used were rat anti-mouse caspase-1 antibody (1:1,000; 4B4; Genentech), rat anti-mouse caspase-11 (1:500; 17D9; Sigma), rabbit anti-IL-1α (1:1,000; ab109555; Abcam), rabbit anti-IL-18 (1:500; 5180R; Biovision), goat anti-mouse IL-1β antibody (1:500; AF-401-NA; R&D Systems) and rabbit anti-GBP2 and rabbit anti-GBP5 (1:1,000; 11854-1-AP/13220-1-AP; Proteintech). Cell lysates were probed with anti-β-actin antibody (Sigma) at 1:2,000.

Statistical analysis. Statistical data analysis was done using Prism 5.0a (GraphPad Software, Inc.). To evaluate the differences between two groups (cell death, cytokine release, FACS, CFU and immunofluorescence-based counts) the two-tailed *t*-test was used. In figures NS indicates 'not significant', *P* values are given in figure legends.

Immunofluorescence. Macrophages were seeded on glass coverslips and infected as described above. At the desired time points cells were washed 3× with PBS and fixed with 4% paraformaldehyde for 15 min at 37 °C. Following fixation coverslips were washed and the fixative was quenched with 0.1 M glycine for 10 min at room temperature. Coverslips were stained with primary antibodies at 4 °C for 16 h, washed 4× with PBS, incubated for 1 h with appropriate secondary antibodies at room temperature (1:500, AlexaFluor, Invitrogen), washed 4× with PBS and mounted on glass slides with Vectashield containing 4′,6-diamidino-2-phenylindole (DAPI) (Vector Labs). Antibodies used were rabbit anti-LC3 (1:1,000; NB600-1384, Novus), mouse anti-LC3 (1:100, 2G6, NanoTools), guinea-pig anti-p62 (1:100, GP62-C, Progen), goat anti-*Salmonella* (1:500, CSA-1 and CSA-1-FITC, KPL), mouse anti-galectin-8 (1:1,000, G5671, Sigma), goat anti-galectin-8 (1:100, AF1305, R&D), rabbit anti-Optineurin (1:100, ab23666, Abcam), rabbit anti-NDP52 (1:100, D01, Abnova), anti-PDI (1:100, ADI-SPA-890, Enzo Lifesciences), anti-Calnexin (1:100, ADI-SPA-860-D, Enzo Lifesciences), goat anti-GBP1-5 (1:100, sc-166960, Santa Cruz Biotech), rabbit anti-GBP2 and rabbit anti-GBP5 (1:100; 11854-1-AP/13220-1-AP; Proteintech). Coverslips were imaged on a Zeiss LSM700 or a Leica SP8 at ×63 magnification. Colocalization studies were performed as blinded experiments, with in general a minimum count of 100 bacteria per coverslip and performed in triplicate. Immunofluorescence based counts of live (mCherry$^+$/FITC$^+$) and dead (mCherry$^-$/FITC$^+$) bacteria were done as blinded experiment on *z* stacks taken from 15 random fields in three biological replicates, with a total of approximately 10,000 bacteria counted.

Immunohistochemistry. Cryosections were blocked in 1% blocking reagent (Invitrogen) and 2% mouse serum (Invitrogen) in TBST (0.05% Tween in 1× TBS pH 7.4), and stained with primary and secondary antibodies (goat anti-CSA1; 1:500; 01-91-99-MG; KPL and anti-GBP2; 1:100; 11854-1-AP; Proteintech). Secondary antibodies included Santa Cruz Biotech sc-362245 and Molecular Probes A21206, A21445 and A21469.

ROS assay. Measurement of oxygen-dependent respiratory burst of BMDMs was performed by chemiluminescence in the presence of 5-amino-2,3-dihydro-1,4-phtalazinedione (luminol, Sigma Aldrich, 66 μM) using a thermostatically (37 °C) controlled luminometer. Both oxygen and nitrogen species were detected ($O_2^{\bullet -}$, $ONOO^-$, $OH^{\bullet}$). Chemiluminescence generation was monitored every minute for 1 h after IFN-γ (100 U ml^{-1}) and/or *Salmonella* challenge and expressed as counts per minute.

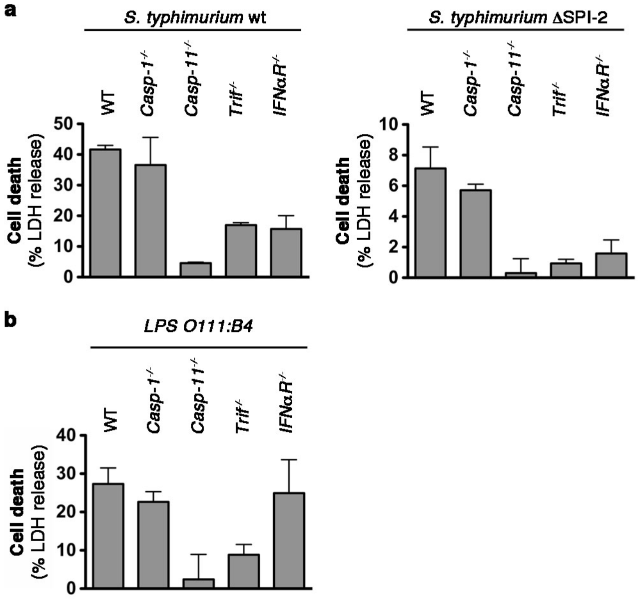
NO assay. Nitrite production was measured by the Griess assay as previously described[26]. Briefly, in 96-well plates, BMDMs were infected as described above in presence or absence of IFN-γ or IL-1β for 16 h. Supernatants were mixed 1:1 with 2.5% phosphoric acid solution containing 1% sulfanilamide and 0.1% naphthylenediamine. After 30 min incubation at room temperature, the nitrite concentration was determined by measuring absorbance at 550 nm. Sodium nitrite (Sigma) was used as a standard to determine nitrite concentrations in the cell-free medium.

Digitonin assay. For flow-cytometry-based quantification of cytoplasmic and vacuolar bacteria, macrophages were infected with mCherry$^+$ *S. typhimurium* or mCherry$^+$ *S. flexneri* as described above. At the desired time point, cells were washed 3× with KHM buffer (110 mM potassium acetate, 20 mM HEPES, 2 mM $MgCl_2$,

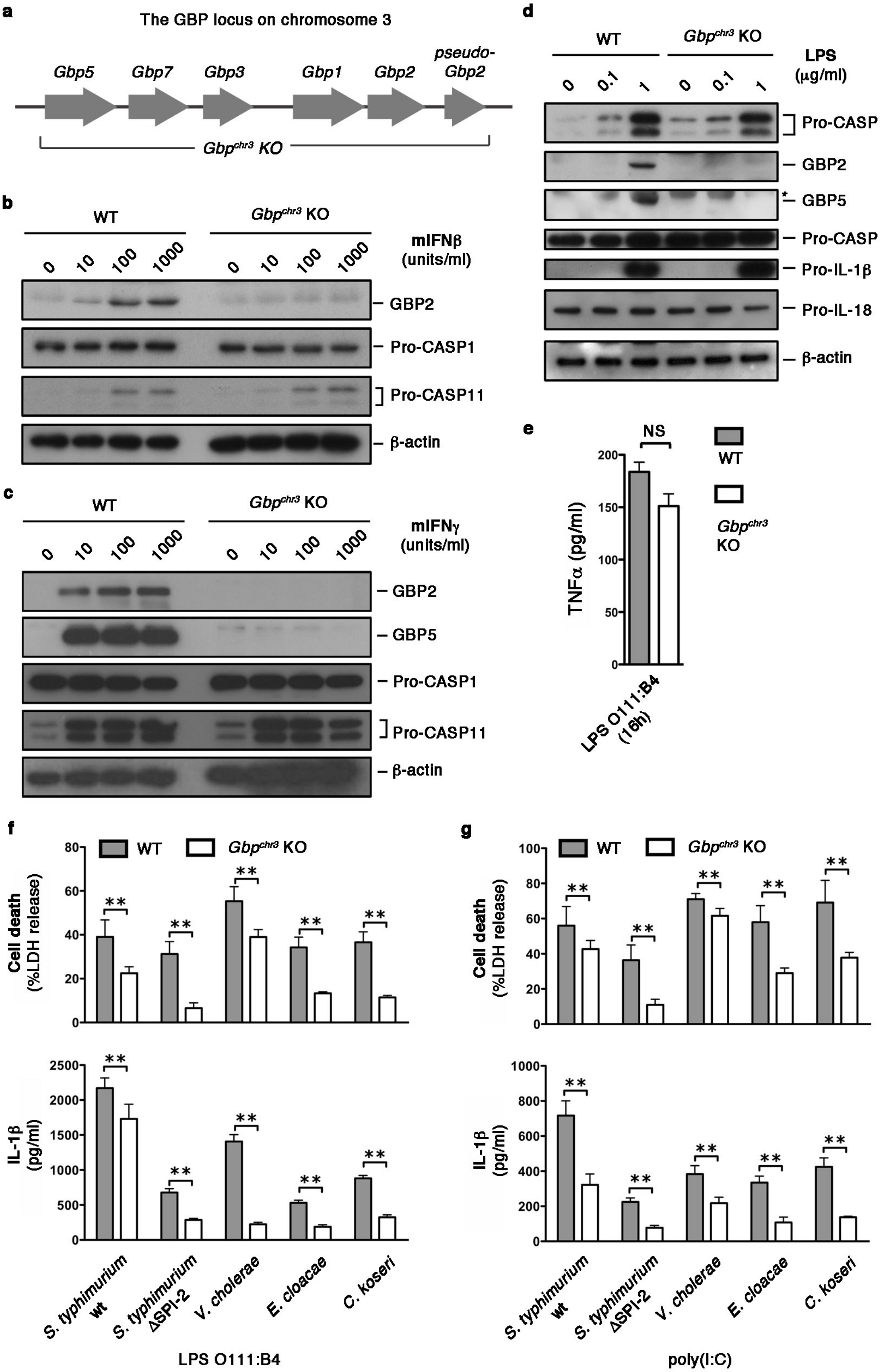
pH 7.3) and incubated for 1 min in KHM buffer with 150 μg ml^{-1} digitonin (Sigma). Cells were immediately washed 2× with KHM buffer and then stained for 12 min with anti-*Salmonella*-FITC (1:500, CSA-1, KPL) or anti-*Shigella* (1:100, BP1064, Acris) in KHM buffer with 2% BSA. Secondary antibodies used for *S. flexneri* staining were: anti-Rabbit-488 (1:500, Invitrogen). Cells were washed 3× with PBS and lysed in PBS with 0.1% Triton-X (Sigma) and analysed on a FACS-Canto-II. Controls were included in every assay and are described in (Extended Data Fig. 9).

Live/dead analysis by FACS. Infection of macrophages was performed using $mCherry^{+}$ bacteria as described above. At 16 h post-infection cells were washed and lysed with PBS solution containing 0.1% Triton X-100 (Sigma Aldrich) to release intracellular bacteria. *Salmonella* were counterstained using an anti-*Salmonella* antibody (CSA-1, KPL) and analysed using a FACS Canto-II for fluorescence intensities in FL-1 and FL-2 channels. Data were analysed with FlowJo 10.0.6 software. The gate was set for the bacterial population based on the FSC/SSC and the anti-*Salmonella* staining (CSA-1-FITC, KPL). Controls included live mCherry-expressing and mCherry-negative *Salmonella* stained with anti-*Salmonella* antibodies (CSA-1, KPL).

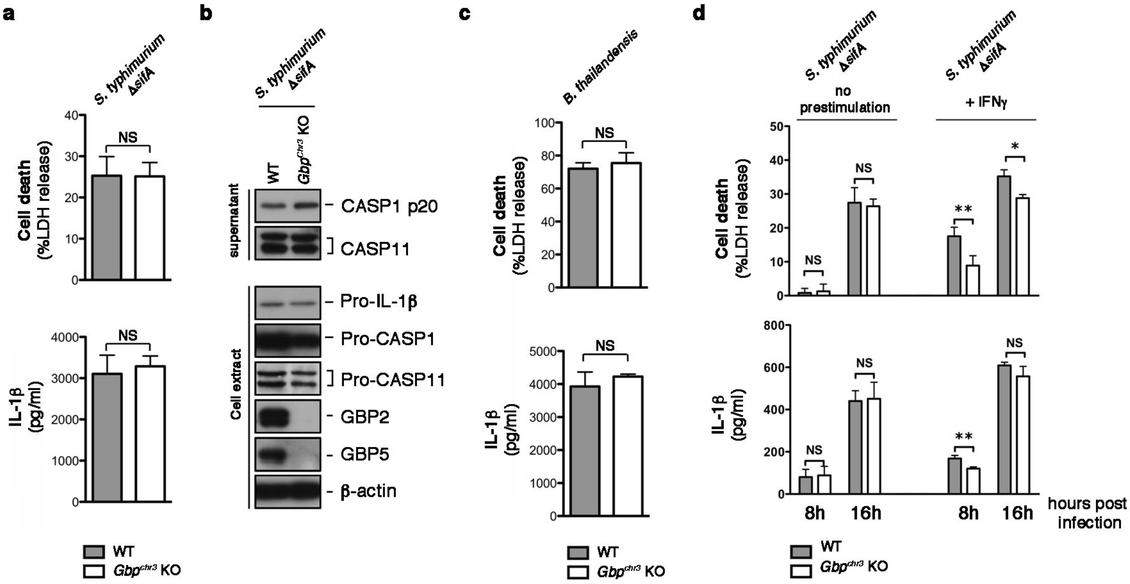
24. Mariathasan, S. *et al.* Differential activation of the inflammasome by caspase-1 adaptors ASC and Ipaf. *Nature* **430,** 213–218 (2004).
25. Zhao, Z. *et al.* Autophagosome-independent essential function for the autophagy protein Atg5 in cellular immunity to intracellular pathogens. *Cell Host Microbe* **4,** 458–469 (2008).
26. Lima-Junior, D. S. *et al.* Inflammasome-derived IL-1β production induces nitric oxide-mediated resistance to *Leishmania*. *Nature Med.* **19,** 909–915 (2013).



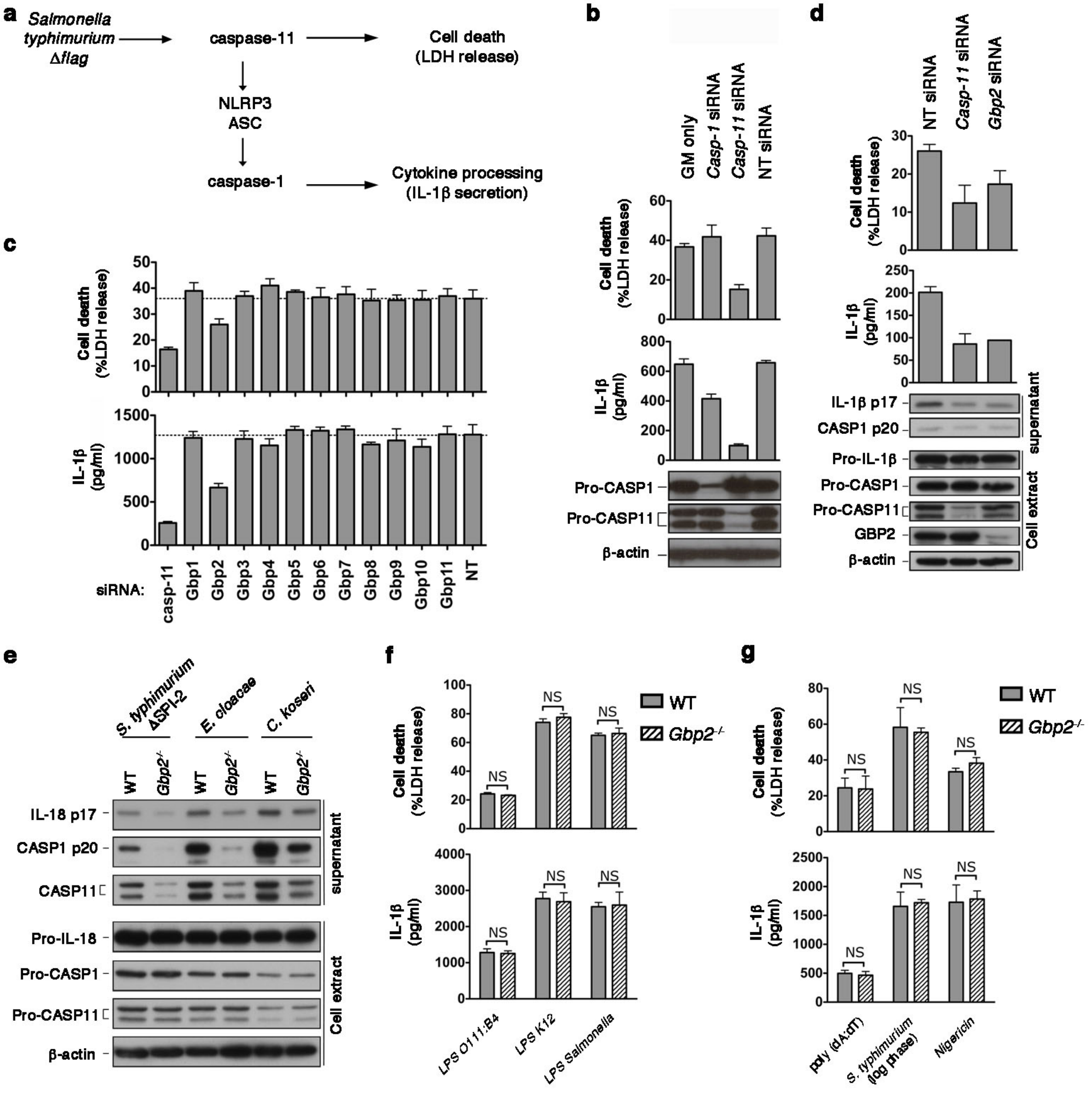
Extended Data Figure 1 | Type-I-interferon signalling is required to induce caspase-11-dependent cell death in response to bacterial infection, but not in response to LPS transfection. **a**, LDH release from unprimed BMDMs infected for 16 h with wild-type (WT) *S. typhimurium* or ΔSPI-2 *S. typhimurium* grown to stationary phase. **b**, LDH release from primed BMDMs transfected with LPS O111:B4. Graphs show the mean and s.d. of quadruplicate wells and are representative of three independent experiments.



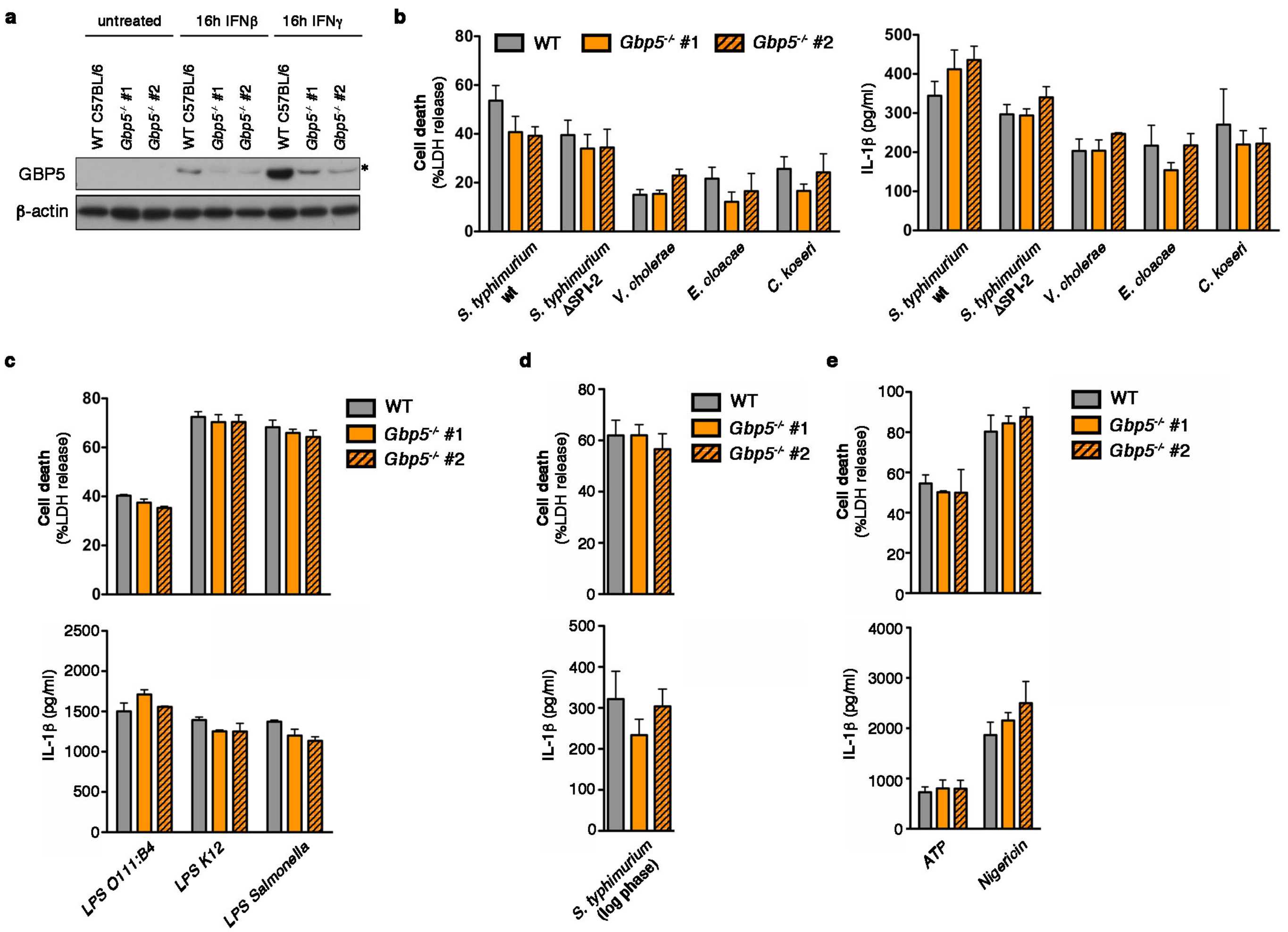
Extended Data Figure 2 | BMDMs from *Gbp^{chr3}* KO mice have normal responses to priming stimuli, but fail to activate the non-canonical inflammasome during bacterial infections. **a**, Schematic representation of the GBP locus on murine chromosome 3. The extent of the deletion in *Gbp^{chr3}* KO mice is indicated. **b**–**d**, Induction of pro-caspase-11, GBP2 and GBP5 expression in lysates of wild-type and *Gbp^{chr3}* KO BMDMs stimulated for 16 h with the indicated amounts of murine IFN-β, murine IFN-γ or LPS O111:B4. **e**, TNF-α release from BMDMs stimulated for 16 h with LPS O111:B4. **f**, **g**, LDH release and IL-1β secretion from wild-type and *Gbp^{chr3}* KO BMDMs infected for 16 h with wild-type (WT) *S. typhimurium*, ΔSPI-2 *S. typhimurium*, *V. cholerae*, *E. cloacae* or *C. koseri* grown to stationary phase. Cells were primed overnight with LPS (**f**) or poly(I:C) (**g**). *Indicates background band. Graphs show the mean and s.d. of quadruplicate wells and data are representative of two independent experiments. $^{**}P < 0.01$, NS, not significant (two-tailed *t*-test).



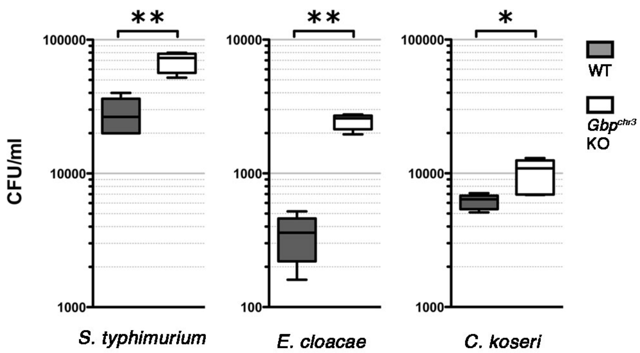
Extended Data Figure 3 | GBPs assist the detection of bacteria that escape into the cytosol only in primed macrophages. **a–c**, LDH release, IL-1β secretion and immunoblots for processed caspase-1 and caspase-11 released from unprimed BMDMs infected for 8–16 h with *ΔsifA S. typhimurium* or *B. thailandensis* grown to stationary phase. **d**, LDH release and IL-1β secretion from unprimed or IFN-γ-primed BMDMs infected for 16 h with *ΔsifA S. typhimurium* grown to stationary phase. Ext, extract; SN, supernatant. Graphs show the mean and s.d. of quadruplicate wells and data are representative of two independent experiments. *$P < 0.05$; **$P < 0.01$; NS, not significant (two-tailed t-test).



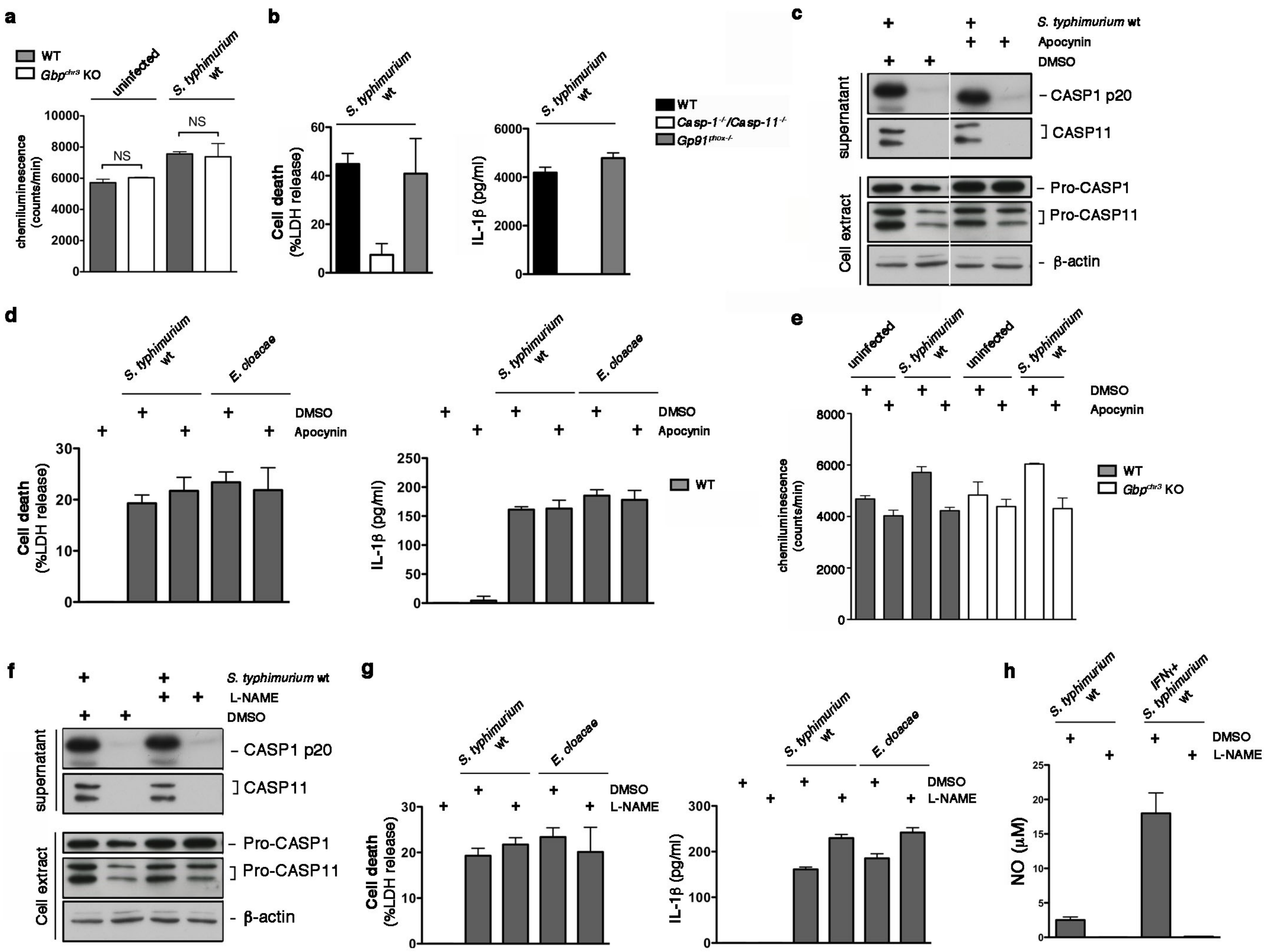
Extended Data Figure 4 | Murine GBP2 controls non-canonical inflammasome activation during *Salmonella* infection, but is dispensable for direct LPS sensing and canonical inflammasomes. **a**, Schematic drawing of the inflammasome pathways activated by flagellin-deficient *Salmonella*. **b**–**d**, LDH release, IL-1β secretion and immunoblots for processed caspase-1 and processed IL-1β released from unprimed BMDMs infected for 17 h with Δflag *S. typhimurium* grown to stationary phase. BMDMs were treated with the indicated siRNA for 56 h before infection. **e**, Immunoblots for processed caspase-1, IL-18 and caspase-11 released from unprimed BMDMs infected for 16 h with ΔSPI-2 *S. typhimurium*, *E. cloacae* or *C. koseri* grown to stationary phase. **f**, **g**, LDH release and IL-1β secretion from primed wild-type and $Gbp2^{-/-}$ BMDMs transfected with the indicated types of LPS for 16 h, treated with nigericin for 1 h, infected with SPI-1 T3SS expressing logarithmic phase wild-type *S. typhimurium* for 1 h, or transfected with poly(dA:dT) for 6 h. Cell were primed with PAM3CSK4 in **f** or LPS **g**. Graphs show the mean and s.d. of quadruplicate wells and data are representative of two (**e**) and three (**b**–**d**, **f**, **g**) independent experiments. NT, non-targeting siRNA; GM, GenMute transfection reagent; NS, not significant (two-tailed *t*-test).



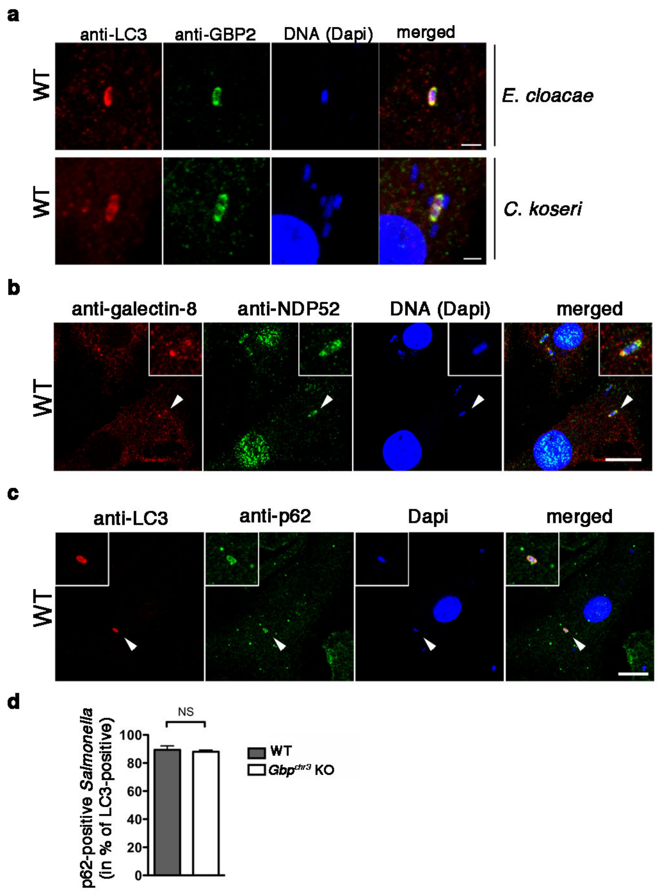
Extended Data Figure 5 | Normal activation of non-canonical and canonical inflammasomes in *Gbp5*$^{-/-}$ BMDMs. **a**, Expression of GBP5 in wild-type and two lines of *Gbp5*$^{-/-}$ BMDMs (1 and 2). *Indicates a cross-reactive band. **b**–**e**, LDH release and IL-1β secretion from BMDMs infected for 16 h with wild-type (WT) *S. typhimurium*, ΔSPI-2 *S. typhimurium*, *V. cholerae*, *E. cloacae* or *C. koseri* grown to stationary phase (**b**), transfected with the indicated LPS for 16 h (**c**) infected for 1 h with SPI-1 T3SS expressing logarithmic phase wild-type *S. typhimurium* (**d**), or treated with 5 mM ATP or 20 mM nigericin for 4 h (**e**). Cells were left unprimed (**b**) or primed with PAM3CSK4 in (**c**) or LPS (**d**, **e**). Graphs show the mean and s.d. of triplicate or quadruplicate wells and data are representative of three independent experiments.



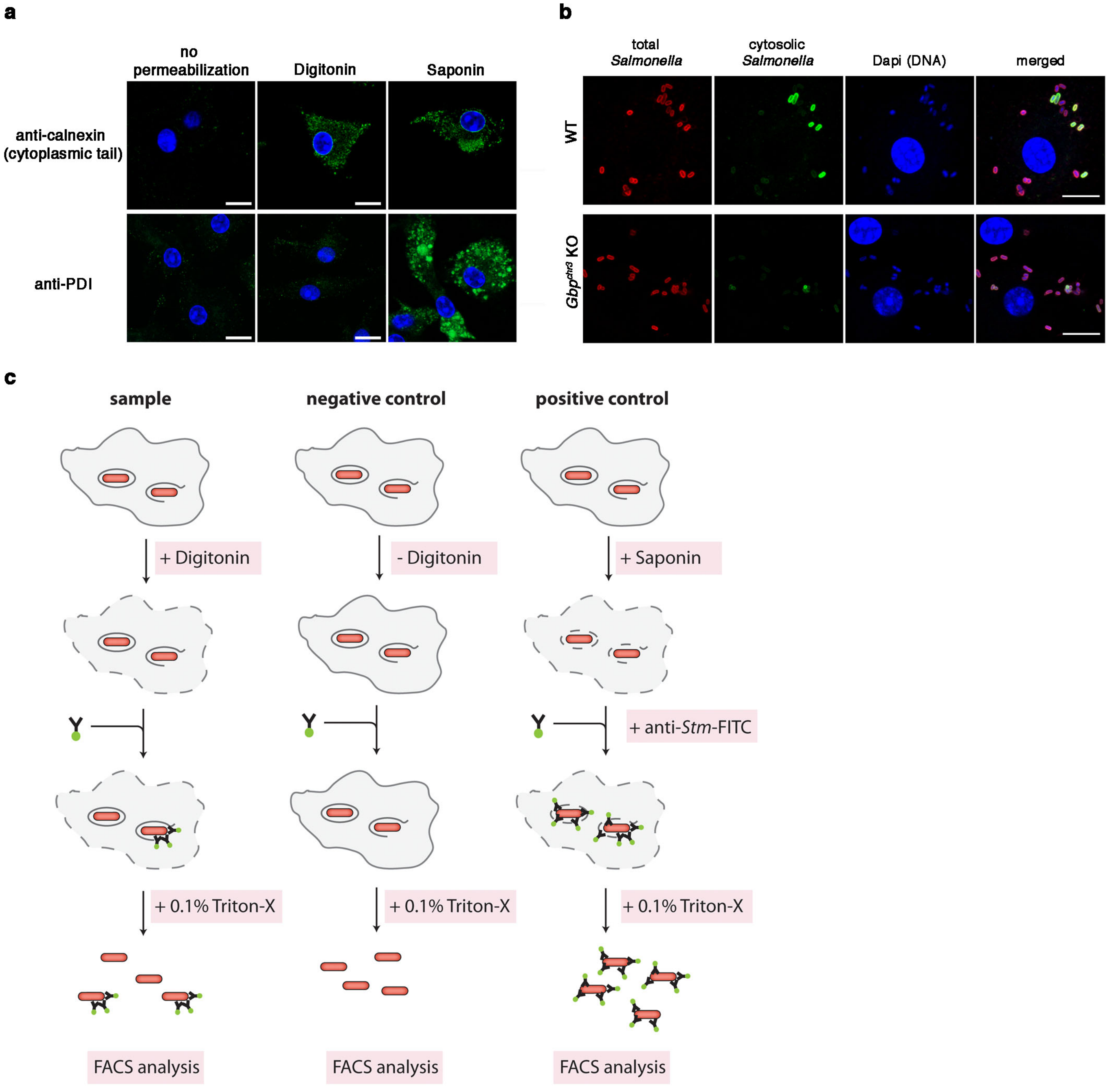
Extended Data Figure 6 | GBPs control bacterial replication. c.f.u.s at 16 h post-infection in wild-type and *Gbp*^{chr3} KO BMDMs infected with the indicated bacterial strains. Experiments are representative of two independent experiments.



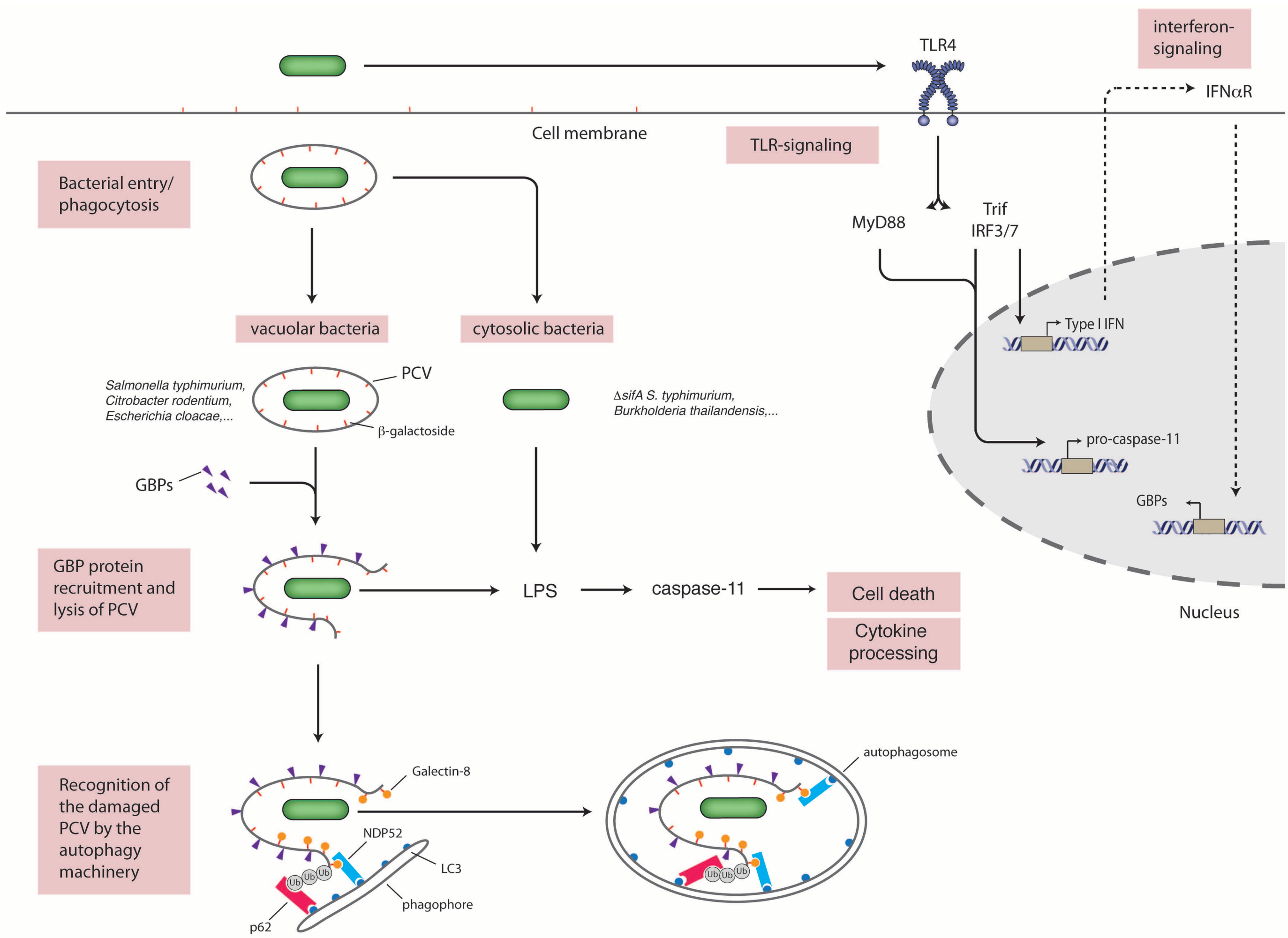
Extended Data Figure 7 | Inhibition of ROS and NO production does not affect non-canonical inflammasome activation. **a**, **b**, ROS levels, LDH release and IL-1β secretion in unprimed BMDMs left uninfected or infected for 16 h with wild-type *S. typhimurium* grown to stationary phase. **c**–**e**, LDH release, IL-1β secretion, ROS levels and immunoblots for processed caspase-1 and caspase-11 released from unprimed BMDMs infected for 16 h with wild-type (WT) *S. typhimurium* or *E. cloacae* grown to stationary phase in the presence of the ROS inhibitor (apocynin) or a vehicle control (DMSO). **f**, **g**, LDH release, IL-1β secretion and immunoblots for processed caspase-1 and caspase-11 released from unprimed BMDMs infected for 16 h with wild-type *S. typhimurium* or *E. cloacae* grown to stationary phase in the presence of the iNOS inhibitor (L-NAME) or a vehicle control (DMSO). **h**, NO release from unprimed or IFN-γ-primed BMDMs infected for 16 h with *S. typhimurium* in presence of the iNOS inhibitor (L-NAME) or a vehicle control (DMSO). Ext, extract; SN, supernatant. Graphs show the mean and s.d. of quadruplicate wells and data are representative of two (**a**–**c**, **e**–**g**) and three (**d**, **h**) independent experiments. NS, not significant (two-tailed *t*-test).



Extended Data Figure 8 | Colocalization of GBPs and autophagy proteins on intracellular bacteria. **a**, Colocalization of LC3 with GBPs in unprimed wild-type BMDMs infected with *E. cloacae* or *C. koseri* for 4 h and stained for LC3, GBP2 and DNA. **b**, Colocalization of galectin-8 and NDP52 in unprimed wild-type BMDMs infected with wild-type *S. typhimurium* for 4 h and stained for galectin-8, NDP52 and DNA. **c**, Colocalization of p62 and LC3 in unprimed wild-type BMDMs infected with wild-type *S. typhimurium* for 4 h and stained for LC3, p62 and DNA. **d**, Quantification of p62 and LC3 co-staining in wild-type and *Gbp*^{chr3} KO BMDMs at 4 h post-infection with *Salmonella*. Arrowheads indicate region shown in insets. Scale bars, 1 μm (**a**) and 10 μm (**b**, **c**). Graph shows the mean and s.d. of triplicate counts and images and graph are representative of at least two independent experiments. NS, not significant (two-tailed *t*-test).



Extended Data Figure 9 | Digitonin-based quantification of cytoplasmic bacteria. **a**, Immunostaining for calnexin and PDI (protein disulphide isomerase) in wild-type BMDMs left untreated or permeabilized with digitonin or saponin. **b**, Differentially permeabilized macrophages stained for cytosolic and vacuolar *Salmonella* at 4 h post-infection. **c**, Schematic representation of FACS-based analysis of cytosolic and vacuolar bacterial populations of *Salmonella*. Scale bars, 10 μm.



Extended Data Figure 10 | Model for the role of GBPs and autophagy in caspase-11 activation. The pathogen-containing vacuole of vacuolar bacterial pathogens is recognized by interferon-induced GBPs in an unknown manner. GBPs promote the lysis of the PCV either directly or indirectly, resulting in the release of the bacteria into the cytosol and activation of caspase-11 by bacterial LPS. β-galactosides of the lysed vacuole serve as danger signals upon exposure to the cytosol and are recognized by galectin-8 leading to the recruitment of the autophagy machinery. p62 participates in this process by recognizing ubiquitin-chains on the vacuole or the bacterium. Uptake of the bacterium and the lysed vacuole into autophagosomes reduces caspase-11 activation by removing the source of LPS from the cytosol.

LETTER

doi:10.1038/nature13173

Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq

Barbara Treutlein[1]*, Doug G. Brownfield[2]*, Angela R. Wu[1], Norma F. Neff[1], Gary L. Mantalas[1], F. Hernan Espinoza[2], Tushar J. Desai[3], Mark A. Krasnow[2] & Stephen R. Quake[1]

The mammalian lung is a highly branched network in which the distal regions of the bronchial tree transform during development into a densely packed honeycomb of alveolar air sacs that mediate gas exchange. Although this transformation has been studied by marker expression analysis and fate-mapping, the mechanisms that control the progression of lung progenitors along distinct lineages into mature alveolar cell types are still incompletely known, in part because of the limited number of lineage markers[1–3] and the effects of ensemble averaging in conventional transcriptome analysis experiments on cell populations[1–5]. Here we show that single-cell transcriptome analysis circumvents these problems and enables direct measurement of the various cell types and hierarchies in the developing lung. We used microfluidic single-cell RNA sequencing (RNA-seq) on 198 individual cells at four different stages encompassing alveolar differentiation to measure the transcriptional states which define the developmental and cellular hierarchy of the distal mouse lung epithelium. We empirically classified cells into distinct groups by using an unbiased genome-wide approach that did not require a priori knowledge of the underlying cell types or the previous purification of cell populations. The results confirmed the basic outlines of the classical model of epithelial cell-type diversity in the distal lung and led to the discovery of many previously unknown cell-type markers, including transcriptional regulators that discriminate between the different populations. We reconstructed the molecular steps during maturation of bipotential progenitors along both alveolar lineages and elucidated the full life cycle of the alveolar type 2 cell lineage. This single-cell genomics approach is applicable to any developing or mature tissue to robustly delineate molecularly distinct cell types, define progenitors and lineage hierarchies, and identify lineage-specific regulatory factors.

In mice, alveolar epithelial cells differentiate between embryonic days (E)16.5 and 18.5: distal airway tips expand into sac-like configurations ('sacculation') as a morphologically uniform population of columnar progenitors proceeds towards the fate of either flat alveolar type 1 (AT1) cells specialized for gas exchange or surfactant-secreting cuboidal alveolar type 2 (AT2) cells (Extended Data Fig. 1). At each time point during sacculation, progenitors, intermediates and recently differentiated cells coexist (Fig. 1a)[6]. To resolve the cellular composition of the developing bronchio-alveolar epithelium, we initially sequenced transcriptomes of 80 individual live cells of the developing mouse lung epithelium late in sacculation (E18.5; three biological replicates). Single-cell suspensions of micro-dissected distal lung regions were purified by magnetic-activated cell sorting (MACS) to deplete leukocytes and alveolar macrophages and enrich for epithelial cells ($CD45^{-}$/ $EpCAM^{+}$) (Extended Data Fig. 2). An automated microfluidic platform was used to capture and lyse individual epithelial cells, reverse transcribe RNA and amplify complementary DNA.

RNA-seq libraries from the amplification products of single cells as well as bulk control samples were sequenced to a depth of $(2–5) \times 10^6$ reads per library (Methods). Saturation analysis confirmed that this sequencing depth is sufficient to detect most genes expressed by single cells (Extended Data Fig. 3a). Technical noise and dynamic range were

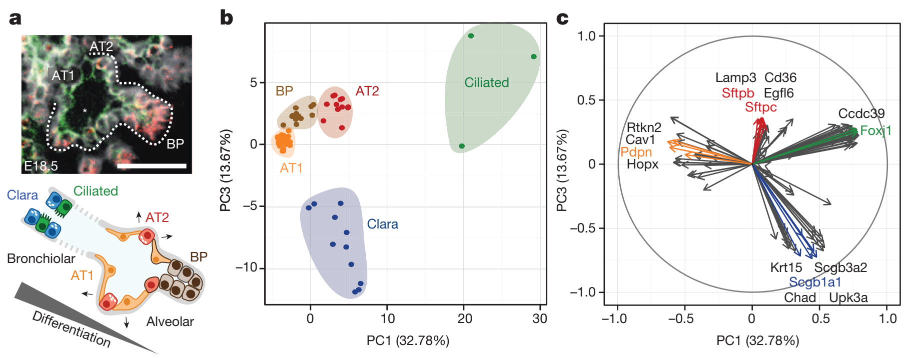


Figure 1 | Single-cell RNA-seq of 80 embryonic (E18.5) mouse lung epithelial cells enables unbiased identification of alveolar, bronchiolar and progenitor cell populations. **a**, Spatially heterogeneous differentiation of distal lung epithelium. The micrograph of a newly forming alveolar sac (asterisk) and the diagram below illustrate cell types and the gradient of developmental intermediates comprising the distal lung epithelium during sacculation (E18.5). Micrograph: green, Pdpn, alveolar type 1 (AT1) marker; red, Sftpc, AT2 marker; white, E-cadherin, pan-epithelial marker. BPs are characterized by co-expression of some AT1 and AT2 markers. In the diagram, BPs (brown) persist at the tip, and nascent AT2 (red) and AT1 (orange) cells are located more proximally. Ciliated (green) and Clara (blue) cells are located in the bronchiolar epithelium (not labelled in the micrograph). Scale bar, 75 μm. **b**, PCA of 80 single-cell transcriptomes (three biological replicates) at E18.5 distinguishes between major bronchiolar and alveolar cell lineages. PC, principal component. **c**, Distinct gene groups characterize each cell population on the basis of differential correlation with PC1 and PC3. The arrow tip denotes the correlation coefficient of the respective gene with each principal component.

[1]Departments of Bioengineering and Applied Physics, Stanford University School of Medicine and Howard Hughes Medical Institute, Stanford, California 94305, USA. [2]Department of Biochemistry, Stanford University School of Medicine and Howard Hughes Medical Institute, Stanford, California 94305, USA. [3]Department of Internal Medicine, Division of Pulmonary and Critical Care Medicine, Stanford University School of Medicine, Stanford, California 94305, USA.
*These authors contributed equally to this work.

assessed by using RNA control spike-in standards and by comparing single cells with the bulk samples (Extended Data Fig. 3b–e). The results are consistent with previous data from our group[7] and others[8–20]; we obtained single-transcript sensitivity and high ($\sim 10^5$) dynamic range. Comparison of three biological replicate experiments showed that median expression of all genes across single cells was strongly correlated ($r = 0.91$ and $r = 0.92$; Extended Data Fig. 3f, g).

We performed principal component analysis (PCA) on all 80 single-cell transcriptomes by using genes expressed in more than two cells and with a non-zero variance (8,578 genes). Genes with highest loadings in the first four principal components were analysed by unsupervised hierarchical clustering as well as PCA (Fig. 1b, c, Fig. 2a and Supplementary Data). This unbiased approach detected five different cell populations and four different gene families, which permutation analysis showed to be highly significant (Methods). Using known marker genes within the different clusters, we were able to associate cells with four previously reported epithelial cell types (Clara (*Scgb1a1*), ciliated (*Foxj1*), AT1 (*Pdpn*, *Ager*) and AT2 (*Sftpc*, *Sftpb*) cells). The fifth group was characterized by co-expression of AT1 and AT2 marker genes and was located on the PCA plot between the populations of AT1 and AT2 cells, suggesting either an intermediate population undergoing a transition between the two alveolar lineages or a population of bipotential alveolar progenitors. As discussed below, transcriptional profiles of distal lung epithelial cells at E16.5 implicate this fifth population as alveolar bipotential progenitor (BP) cells[6]. We validated these findings in two biological replicates of pooled E18.5 lungs by microfluidic single-cell quantitative PCR (qPCR) experiments: hierarchical clustering of ten known alveolar and bronchiolar marker genes identified the same five populations (Extended Data Fig. 5a–d). Together, these results show that single-cell RNA-seq enables the identification and molecular characterization of cell types and developmental intermediates retrospectively without the need to first purify populations of interest.

In addition to classifying the epithelial cell populations in the distal lung at E18.5, our analysis identified sets of genes specific to each population, providing a battery of previously unknown markers that can be used to distinguish cells from each alveolar and bronchiolar lineage. We used Guilt-by-Association and correlation analysis to assess the significance of co-expression of genes in all cells belonging to a specific cell type (Methods, Fig. 2b and Supplementary Data). The large number of lineage-specific genes allowed us to annotate functions of individual cell types by gene ontology and pathway enrichment analysis[21] (Extended Data Fig. 4a and Supplementary Data): AT1 cells were enriched in pathways associated with extracellular matrix-receptor interaction, focal adhesion, tight and adherens junctions and regulation of the actin cytoskeleton; AT2 cells were enriched for adipocytokine and PPAR signalling and for lysosome pathways; the Clara cell lineage was enriched for metabolism of xenobiotics by cytochrome P450, drug metabolism and glutathione metabolism; and ciliated cells showed enrichment for progesterone-mediated oocyte maturation and cell cycle pathways. Furthermore, we identified transcription factors, receptors and ligands whose expression profile across all single cells was strongly correlated with the individual cell types (Extended Data Fig. 4b, c).

Among the numerous newly identified putative cell-type markers, several are of particular interest. *Hopx* transcription factor was previously reported to regulate alveolar maturation by suppressing surfactant protein production in AT2 cells[22]; our data show that *Hopx* is expressed in BPs, turns off in maturing AT2 cells and is maintained in AT1 cells. We validated the AT1-specific expression of *Hopx* by transgenic labelling and co-localization with two AT1 markers, *Pdpn* and *Ager* (Fig. 2c and Extended Data Fig. 4e). We also found that *Vegfa* endothelial growth factor is specifically expressed in the AT1 lineage, presumably serving as a signal to activate nearby capillary endothelial cells; AT1-specific expression was validated by single-cell qPCR (Extended Data Fig. 4d). *Egfl6*, encoding a protein implicated in cell adhesion and cell differentiation, is specifically expressed in AT2 cells; AT2-specific expression was confirmed by multiplex *in situ* hybridization with the canonical AT2

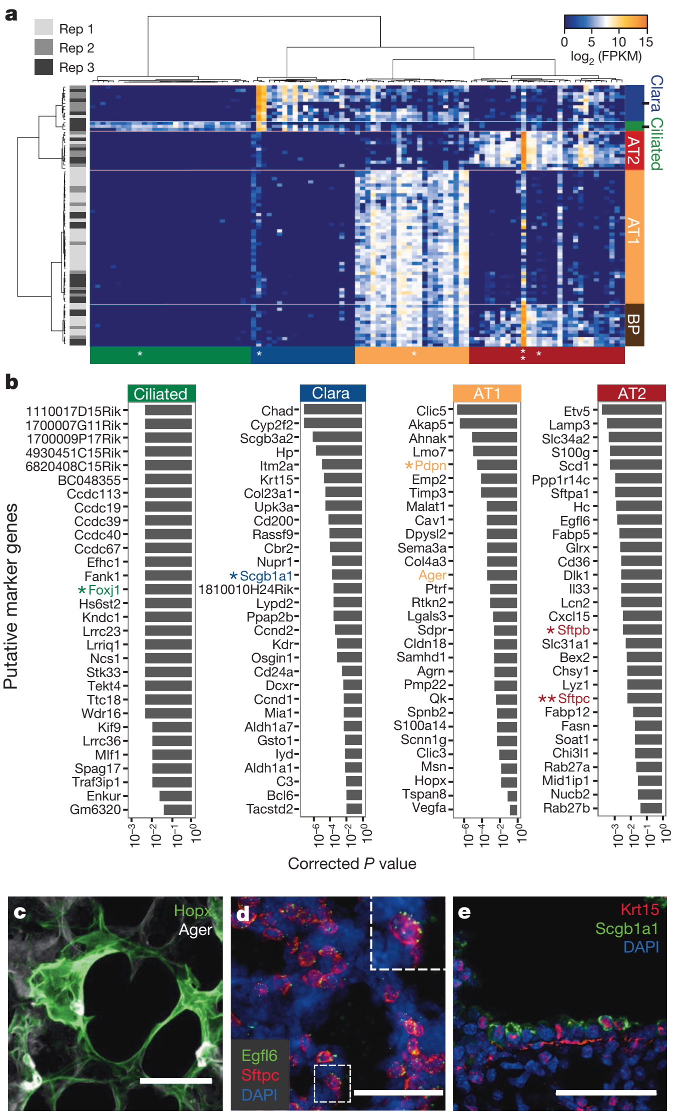


Figure 2 | Single-cell transcriptome analysis discovers previously unknown markers. a, Hierarchical clustering of RNA-seq data from 80 single distal lung epithelial cells (E18.5, three biological replicates) identifies five molecularly distinct populations, assigned to alveolar and bronchiolar lineages based on the presence of canonical marker genes (asterisks) within the respective gene clusters (AT2 (red), *Sftpb* and *Sftpc*; AT1 (orange). *Pdpn*; ciliated (green), *Foxj1*; Clara (blue), *Scgb1a1*). BPs (brown) co-express AT1 and AT2 markers. Each row represents a single cell, each column a gene (104 genes in total; Supplementary Data). Permutation analysis supports the significance of the presented clustering ($P = 2.89 \times 10^{-122}$; Methods). FPKM, fragments per kilobase of transcript per 10^6 mapped reads. **b**, Bar graphs showing the top 30 putative marker genes for each cell lineage inferred from the E18.5 single-cell transcriptomes as a function of the multiple testing corrected *P* value for each gene (Guilt-by-Association; Methods). Canonical markers are bold and coloured. **c**, Validation of *Hopx* expression in AT1 cells. A lung section from a transgenic *Hopx-Cre-ERT2*$^{+/-}$;*mTmG*$^{+/tg}$ adult mouse was co-stained for AT1 marker Ager. Maximum intensity projections of confocal *z* stacks show that AT1 cells expressing membrane-localized green fluorescent protein (GFP; green) also express Ager (white). Scale bar, 50 μm. **d**, Validation of *Egfl6* expression in AT2 cells. Multiplexed *in situ* hybridization of E18.5 lungs shows co-localization of probes targeting *Egfl6* (green) and AT2 marker *Sftpc* (red) messenger RNA. Inset, close-up of boxed region. Blue, 4′,6-diamidino-2-phenylindole (DAPI)-stained nuclei. Scale bar, 50 μm. **e**, Validation of Krt15 expression in Clara cells. Immunofluorescent staining of E18.5 lungs with the use of antibodies against Krt15 (red) and Clara cell marker Scgb1a1 (green). Blue, DAPI-stained nuclei. Krt15 is also expressed outside the epithelium. Scale bar, 50 μm.

marker *Sftpc* (Fig. 2d). Krt15, a component of intermediate filaments, was specifically expressed in the Clara cell lineage, which we validated by co-staining with the canonical Clara cell marker Scgb1a1 (Fig. 2e). Finally, we used single-cell multiplexed qPCR to validate the lineage-specific expression of six additional genes including *Itgb4* and *Top2a* for ciliated cells, *Cftr*, *Cebpa*, *Sftpd* and *Id2* for the AT2 lineage, and *Vegfa* for the AT1 lineage (Extended Data Fig. 4d). Most genes specifically expressed by the AT2 lineage at E18.5 were also detected by single-cell RNA-seq in mature AT2 cells of an adult mouse lung, whereas genes specific to AT1, Clara or ciliated cells were not expressed or were expressed only at a low level (Extended Data Fig. 4f). Thus, we identified a large number of new, and potentially more specific, markers for various biological processes and stages relevant to alveolar and bronchiolar maturation.

Identification of the progenitor and differentiated cell types at E18.5 prompted further investigation of developmental intermediates in the alveolar maturation pathway. Sacculation of distal airway tubules commences at E16.5, and the distal epithelium is dominated by alveolar progenitor cells at this time[6]. We therefore measured transcript levels of ten known marker genes in 107 single cells of the distal lung epithelium at both E16.5 (33 cells) and E18.5 (74 cells) with multiplexed single-cell qPCR (Extended Data Fig. 5a–d). The marker gene expression profile and PCA identified Clara and ciliated cells distinct from alveolar lineages at both E16.5 and E18.5, corroborating the earlier separation of bronchiolar from alveolar maturation pathways. However, gene expression of alveolar cells showed no segregation into AT1 and AT2 lineages at E16.5, because marker genes for both subpopulations were commonly expressed by all cells, whereas by E18.5 they had clearly separated. This is consistent with a recent temporo-spatial marker study suggesting that AT1 and AT2 lineages emerge from a common BP[6]. In addition to BPs and mature alveolar cells at E18.5, we observed cells in intermediate maturation stages on the basis of partial co-expression of AT1 and AT2 marker genes. We used the newly identified genes specific for each mature alveolar cell type to subclassify these intermediates and thereby reconstruct the molecular pathway of differentiation of BPs into AT1 and AT2 lineages, grouping the genes into early and late markers of either lineage (Fig. 3). We confirmed the presence of developmental intermediates showing heterogeneity in marker gene expression by immunofluorescence (Extended Data Fig. 5f–i).

The constructed hierarchy identified transcription factors, receptors and ligands showing expression changes that correlated with specific transitions in the maturation states of alveolar cells (Extended Data Fig. 5e). The transcription factors *Sox9* and *Cited2* were expressed in BP and AT2 cells, whereas *Hes1* was expressed in BP and AT1 cells. We did not detect any transcription factors that initiated expression exclusively in either of the maturing alveolar lineages, suggesting that lineage commitment involves the downregulation of factors active in alveolar progenitor cells rather than the *de novo* expression of a lineage-specific transcription factor. Ligands were expressed in either BP and AT2 (*Cxcl15*, *Cmtm8*) or BP and AT1 cells (*Sema3a*, *Tgfb*, *Vegfa*), and receptors were expressed in a BP (*Fzd2*), BP/AT2 (*Fgfr2*) or BP/AT1 (*Gprc5a*) pattern. These results show that our approach can be used to characterize transcriptional profiles of transient cellular intermediates during a dynamic maturation process within a complex tissue.

Finally, we explored temporal changes within the distal lung by sequencing additional single-cell transcriptomes before sacculation (E14.5; 45 progenitor cells), early in sacculation (E16.5; 27 progenitor cells) and long after sacculation (adult; 46 transgenically labelled AT2 cells). We performed unsupervised hierarchical clustering analysis of *Sftpc*-positive cells (124 cells), using genes with the highest principal-component loadings in a PCA analysis (Fig. 4a, b and Supplementary Data). Cells clustered in groups that were highly correlated with developmental stage of cell isolation, in sequence from early progenitors (EPs), through BP and nascent AT1 and AT2 cells, to mature AT2 cells. Thus, AT2 cell maturation occurs in a progressive manner through transcriptionally distinct intermediates that can be robustly discriminated by expression

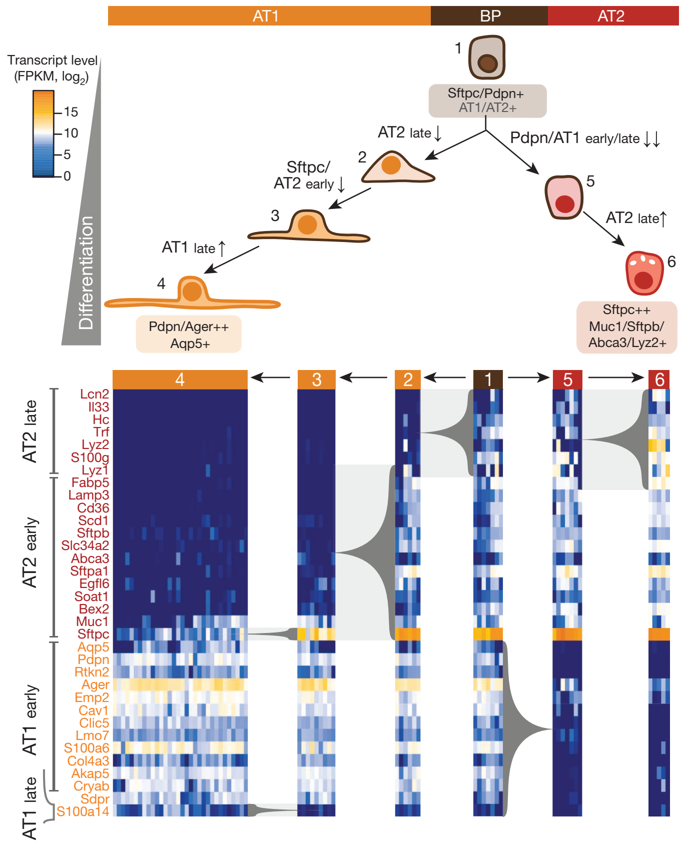


Figure 3 | Molecular profiles distinguish between developmental intermediates during the differentiation of AT1 and AT2 cells from a common BP. Developmental sequence of AT1 (orange) and AT2 (red) specification from a common BP (brown). Two and three maturation intermediates were identified in the specification processes of AT2 and AT1 cell types, respectively, on the basis of the expression of known and previously unknown marker genes for both alveolar lineages measured by single-cell RNA-seq. Genes were grouped into early and late markers of each lineage. Arrows, differentiation pathway; grey braces, change in transcript level of respective genes, tip pointing towards lower expression.

profile throughout embryonic and adult life. The population of EP cells co-expresses the AT2 marker *Sftpc* and the AT1 marker *Pdpn*, indicating that these cells are located at the tips of the branching epithelial tree (Extended Data Fig. 6a). EP cells segregate into two subgroups, one exclusive to E14.5 (early EPs; EP-A) and the other present at both E14.5 and E16.5 (late EPs; EP-B), indicating that cellular differentiation is not fully synchronous throughout the lung. Both EP populations show high expression of genes involved in cell cycle progression and chromosome dynamics (gene groups IIIa, IIb and IIIa; Fig. 4a, c), which are downregulated during the transition of EPs to BPs. The downregulated EP-specific genes include transcription factor *Sox11*, which is expressed in the developing airway epithelium and causes an alveolar defect when knocked out[23], and also *Tuba1a*, a putative target of *Sox11*[24]; this suggests that *Sox11* could be involved in maintaining the proliferative competence of EP cells. At E18.5, BP cells expressing both AT1 and AT2 markers appear in conjunction with intermediate populations with a decreased expression of AT1 markers (nascent AT2) or AT2 markers (nascent AT1). Mature AT2 cells are characterized by the expression of genes involved in respiratory gas exchange and immune response (gene group IV) and were only detected at adult stages (Fig. 4a). The overall number of genes as well as the total number of transcripts expressed in each cell were strongly correlated with its differentiation state: early progenitor cells at E14.5 expressed up to 6,000 genes, whereas mature AT2 cells expressed about fourfold to sixfold fewer genes (Fig. 4a and Extended Data Fig. 7a). Thus, we followed the full life cycle of *Sftpc*$^{+}$ cells and identified seven gene sets that robustly distinguish between

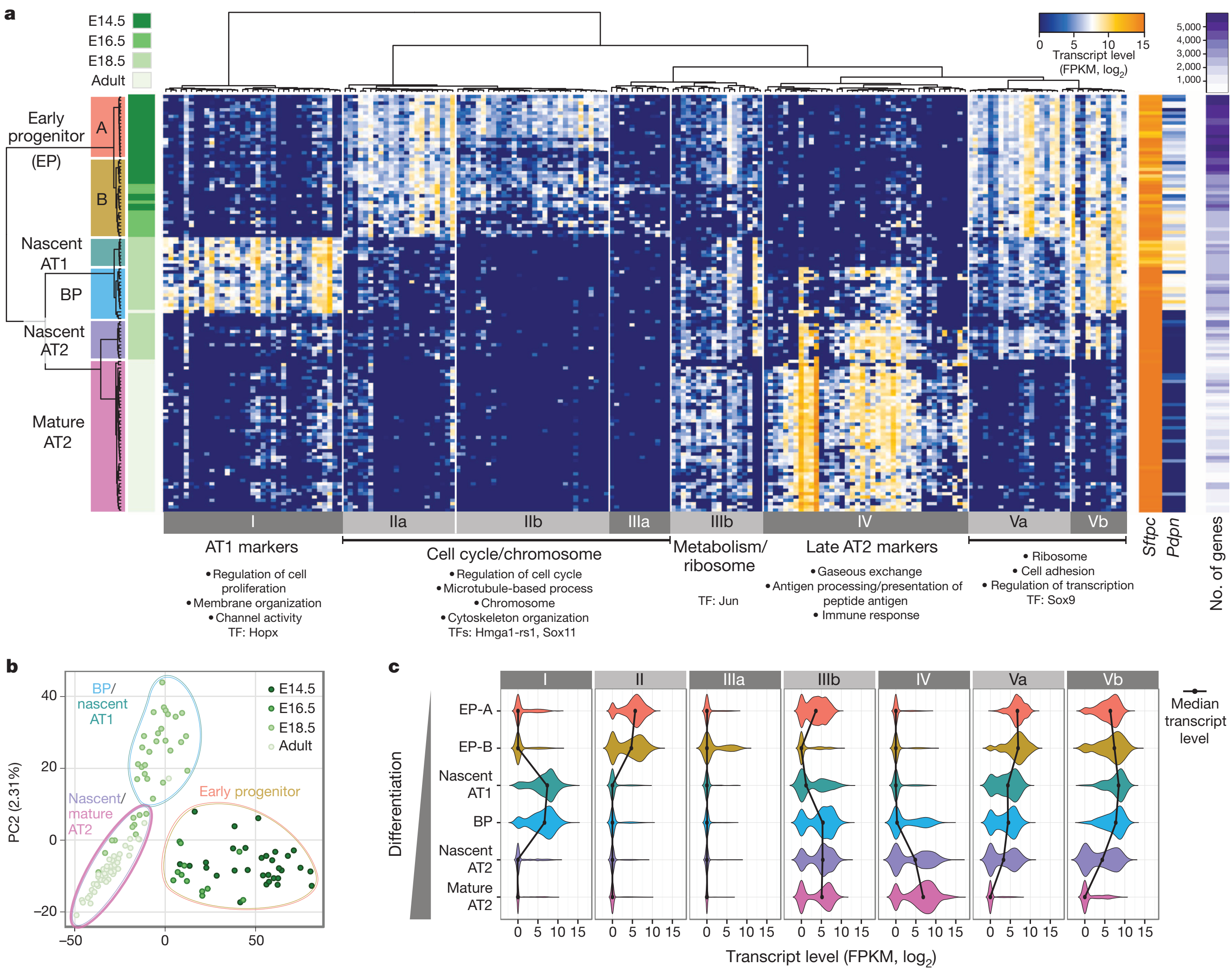


Figure 4 | Single-cell RNA-seq of *Sftpc*$^+$ cells at E14.5, E16.5, E18.5 and in the adult mouse lung explains progressive transcriptional states of the AT2 cell lineage throughout its life cycle. a, Hierarchical clustering of 124 *Sftpc*$^+$ cells from distal mouse lung epithelium of embryonic (E14.5, E16.5 and E18.5) and adult mice based on genes with highest principal-component loadings (Supplementary Data) in an unbiased PCA analysis (shown in **b**) of all cells and genes. Single cells are shown in rows; genes are shown in columns. Bars at the right show *Sftpc* and *Pdpn* expression, as well as the number of genes expressed by each single cell (see also Extended Data Fig. 7). Functional gene ontology enrichments and transcription factors (TFs) specific to each gene group (bottom grey-shaded bars) are shown (Supplementary Data). A similar analysis following the Clara cell lineage throughout development is shown in Extended Data Fig. 8. **b**, PCA of single-cell transcriptomes based on genes detected in more than two cells. Cells cluster into three major populations on the basis of different scores along the first two principal components. **c**, Violin plots depicting the course of expression of each of seven distinct gene groups across the six cell populations. Each violin plot shows the frequency distribution of the mean transcript level ($\log_2$-transformed FPKM) of all genes per cell.

multipotential, bipotential, nascent and mature AT2 cell states (Extended Data Fig. 6b).

We expect that a similar strategy to that pursued here can be applied to almost any tissue to empirically classify and characterize the full set of developing and mature cell types, explain the molecular regulation of these distinct populations, and explore how they are disrupted in disease.

METHODS SUMMARY

Single-cell suspensions were prepared from micro-dissected distal regions of embryonic mouse lung at E14.5, E16.5 and E18.5 and also from adult mouse lung. Epithelial cells were purified by either magnetic bead-activated cell sorting (MACS; Miltenyi Biotec) using CD45 depletion and EpCAM enrichment or by fluorescence-activated cell sorting (FACS) of transgenically labelled cells. Single cells were captured on a microfluidic chip on the C1 system (Fluidigm) and whole-transcriptome amplified cDNA was prepared on chip using the SMARTer Ultra Low RNA kit for Illumina (Clontech). Single-cell libraries were constructed as described previously[7] with the use of the Illumina Nextera XT DNA Sample Preparation kit and sequenced to a depth of $(2–5) \times 10^6$ read pairs (HiSeq 2000; Illumina). Expression levels of transcripts were quantified with TopHat/Cufflinks[25]. Single-cell RNA-seq data was validated by the preparation of single-cell qPCR amplicons of 96 genes on a C1 microfluidic chip (Fluidigm) followed by multiplexed qPCR with BioMark (Fluidigm) as described previously[26]. Single-cell gene expression data was analysed with custom R scripts[27] (Supplementary Data). Gene ontology enrichment analysis was performed with DAVID informatics resources 6.7 (ref. 21). Additional experiments were performed with immunofluorescence and *in situ* hybridization (ACD's RNAscope *In situ* Hybridization Technology).

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 23 July 2013; accepted 19 February 2014.

Published online 13 April 2014.

1. Kim, C. F. B. *et al.* Identification of bronchioalveolar stem cells in normal lung and lung cancer. *Cell* **121,** 823–835 (2005).

2. Zemke, A. C. *et al.* Molecular staging of epithelial maturation using secretory cell-specific genes as markers. *Am. J. Respir. Cell Mol. Biol.* **40,** 340–348 (2009).
3. Guha, A. *et al.* Neuroepithelial body microenvironment is a niche for a distinct subset of Clara-like precursors in the developing airways. *Proc. Natl Acad. Sci. USA* **109,** 12592–12597 (2012).
4. Gonzalez, R. *et al.* Freshly isolated rat alveolar type I cells, type II cells, and cultured type II cells have distinct molecular phenotypes. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **288,** L179–L189 (2005).
5. Xu, Y. *et al.* Transcriptional programs controlling perinatal lung maturation. *PLoS ONE* **7,** e37046 (2012).
6. Desai, T. J., Brownfield, D. G. & Krasnow, M. A. Alveolar progenitor and stem cells in lung development, renewal and cancer. *Nature* **507,** 190–194 (2014).
7. Wu, A. R. *et al.* Quantitative assessment of single-cell RNA-sequencing methods. *Nature Methods* **11,** 41–46 (2013).
8. Islam, S. *et al.* Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Res.* **21,** 1160–1167 (2011).
9. Islam, S. *et al.* Highly multiplexed and strand-specific single-cell RNA 5′ end sequencing. *Nature Protocols* **7,** 813–828 (2012).
10. Shalek, A. K. *et al.* Single-cell transcriptomics reveals bimodality inexpression and splicing in immune cells. *Nature* **498,** 236–240 (2013).
11. Sasagawa, Y. *et al.* Quartz-Seq: a highly reproducible and sensitive single-cell RNA-Seq reveals non-genetic gene expression heterogeneity. *Genome Biol.* **14,** R31 (2013).
12. Liu, C. L., Bernstein, B. E. & Schreiber, S. L. Whole genome amplification by T7-based linear amplification of DNA (TLAD). II. Second-strand synthesis and in vitro transcription. *CSH Protocols*, http://dx.doi.org/10.1101/pdb.prot5003 (2008).
13. Hashimshony, T., Wagner, F., Sher, N. & Yanai, I. CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell Rep.* **2,** 666–673 (2012).
14. Ramskold, D. *et al.* Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nature Biotechnol.* **30,** 777–782 (2012).
15. Tariq, M. A., Kim, H. J., Jejelowo, O. & Pourmand, N. Whole-transcriptome RNAseq analysis from minute amount of total RNA. *Nucleic Acids Res.* **39,** e120 (2011).
16. Picelli, S. *et al.* Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nature Methods* **10,** 1096–1098 (2013).
17. Brennecke, P. *et al.* Accounting for technical noise in single-cell RNA-seq experiments. *Nature Methods* **10,** 1093–1095 (2013).
18. Tang, F. *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. *Nature Methods* **6,** 377–382 (2009).
19. Tang, F. *et al.* RNA-Seq analysis to capture the transcriptome landscape of a single cell. *Nature Protocols* **5,** 516–535 (2010).
20. Tang, F. *et al.* Tracing the derivation of embryonic stem cells from the inner cell mass by single-cell RNA-Seq analysis. *Cell Stem Cell* **6,** 468–478 (2010).
21. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4,** 44–57 (2009).
22. Yin, Z. *et al.* Hop functions downstream of Nkx2.1 and GATA6 to mediate HDAC-dependent negative regulation of pulmonary gene expression. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **291,** L191–L199 (2006).
23. Sock, E. *et al.* Gene targeting reveals a widespread role for the high-mobility-group transcription factor Sox11 in tissue remodeling. *Mol. Cell. Biol.* **24,** 6635–6644 (2004).
24. Wang, X. *et al.* Gene expression profiling and chromatin immunoprecipitation identify DBN1, SETMAR and HIG2 as direct targets of SOX11 in mantle cell lymphoma. *PLoS ONE* **5,** e14085 (2010).
25. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25,** 1105–1111 (2009).
26. Dalerba, P. *et al.* Single-cell dissection of transcriptional heterogeneity in human colon tumors. *Nature Biotechnol.* **29,** 1120–1127 (2011).
27. R core team. *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing* http://www.R-project.org/.

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank W. Koh and B. Passarelli for help and discussions regarding bioinformatic pipelines and statistical analysis, S. I. Gonzalez for help with immunofluorescence, and J. G. Camp and members of the Krasnow laboratory for critical discussion and reading of the manuscript. This work was supported by National Heart, Lung, and Blood Institute (NHLBI) U01HL099995 Progenitor Cell Biology Consortium Grant (B.T., M.A.K., S.R.Q.), by National Institutes of Health (NIH) T32HD007249 (D.G.B.), by a Parker B. Francis Foundation Fellowship and NIH 5K08HL084095 Award (T.J.D.), and by NIH grant U01HL099999 (A.R.W., N.F.N.). M.A.K. and S.R.Q. are investigators of the Howard Hughes Medical Institute.

Author Contributions B.T., D.G.B., T.D., M.A.K. and S.R.Q. conceived the study and designed the experiments. B.T., D.G.B., F.H.E., A.R.W., N.F.N., G.L.M. and T.D. performed the experiments. B.T., D.G.B., A.R.W., F.H.E., T.D., M.A.K. and S.R.Q. analysed the data and/or provided intellectual guidance in their interpretation. B.T., D.G.B., F.H.E., T.D., M.K. and S.R.Q. wrote the paper.

Author Information The transcriptome sequencing data for all single cells has been deposited in the Gene Expression Omnibus database under accession number GSE52583. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the paper on www.nature.com/nature. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.R.Q. (quake@stanford.edu), M.A.K. (krasnow@stanford.edu), T.J.D. (tdesai@stanford.edu).

METHODS

Mouse strains. Timed-pregnant C57BL/6J females (JAX) were used for all embryonic time points reported; gestation age was verified by crown–rump length before use. For adult mice, a transgenic-labelling approach was employed to enrich for AT2 cells. Mice were bred to be homozygous for a knock-in allele into *Sftpc* encoding for a reverse tetracycline transactivator (SftpC-Cre-ERT2-rtTA)[28] and heterozygous for an inserted transgene that drives the expression of a GFP tagged human histone 1 in a tetracycline-dependent manner (tetO-HIST1H2BJ/GFP). To validate the expression of *Hopx*, mice were bred to be hetereozygous for the knock-in of a tamoxifen inducible Cre recombinase (Cre-ERT2) construct into the *Hopx* gene (Hopx-Cre-ERT2)[29] and heterozygous for a transgenic insertion into the *Rosa26* locus encoding a two-colour, membrane-tethered fluorophore reporter that switches expression from a red to green fluorophore on Cre-mediated recombination (mTmG) (cross of B6;129S-$Hopx^{tm1Eno}$/J and B6.129(Cg)-$Gt(ROSA)26Sor^{tm4(ACTB\text{-}tdTomato,\text{-}EGFP)Luo}$/J). Genotyping was performed using PCR with published primer sets from genomic DNA extracted from tails by Proteinase K (Sigma) digestion and precipitation with ethanol. Mice were housed in filtered cages and all experiments were performed in accordance with approved Institutional Animal Care and Use Committee protocols.

Isolation and disaggregation of lung tissue. Single-cell experiments were performed on embryonic mouse lung at E14.5, E16.5 and E18.5 and also on adult mouse lung. In general, embryonic experiments were performed on pooled sibling lungs of one litter (five to seven lungs per pool). One of three replicate experiments at E18.5 (cells referred to as 'E18_2_Cxx' in Supplementary Data) was performed on a single embryonic lung.

Adult mice were euthanized by administration of CO_2. For time points E14.5, E16.5 and E18.5, embryos were removed and lungs were isolated *en bloc* without perfusion and pooled by litter (five to seven embryos) for further processing. Lungs from E14.5 and E16.5 time points were dissociated in Dispase (BD Biosciences) and triturated with glass Pasteur pipettes until a single-cell suspension was attained. For E18.5 and adult time points, either total lung (adult) or peripheral lobe edges (E18.5) were minced with a razor blade into 1 mm^3 fragments, suspended in 5 ml of digestion buffer consisting of elastase (3 U ml^{-1}; Worthington Biochemical Corporation) and DNase I (0.33 U ml^{-1}; Roche) in DMEM/F12 medium, incubated with frequent agitation at 37 °C for 45 min, and triturated briefly with a 5-ml pipette. For all time points, an equal volume of DMEM/F12 supplemented with 10% FBS and penicillin–streptomycin (1 U ml^{-1}; Thermo Scientific) was added to single-cell suspensions before passing the suspension through a 100-µm mesh filter (Fisher) and centrifugation at 400*g* for 10 min. Pelleted cells were resuspended in red blood cell lysis buffer (BD Biosciences), incubated for 2 min, passed through a 40-µm mesh filter (Fisher), centrifuged at 400*g* for 10 min and then resuspended in sorting buffer (PBS supplemented with 0.05% BSA and 2 mM EDTA).

Purification of embryonic distal lung epithelial cells by MACS. Lung epithelial cells for embryonic time points (E14.5, E16.5 and E18.5) were purified by MACS using MS columns (Miltenyi Biotec) in MACS buffer (2 mM EDTA, 0.5% BSA in PBS, filtered and degassed) in accordance with the protocol provided by the vendor. Before loading, the single-cell suspension was passed through a 35-µm cell strainer (BD Biosciences). Leukocytes and alveolar macrophages were removed by depletion with an antibody against the surface antigen CD45 conjugated to magnetic beads (Miltenyi Biotec) followed by enrichment for epithelial cells, incubating first with a biotinylated primary antibody for EpCAM (clone G8.8; eBioscience) followed by a secondary antibody against biotin conjugated to magnetic beads (Miltenyi Biotec).

Purification of adult AT2 cells by FACS. For AT2 cells from the adult lung, an adult Sftpc-Cre-ERT2-rtta$^{-/-}$ tetO-HIST1H2BJ-GFP$^{+/-}$ mouse was injected with 2 mg of doxycycline (Sigma) three days prior to the single cell experiments. After isolation and disaggregation of the lung, the single-cell suspension was incubated with a viability stain (Sytox Blue; Invitrogen) for 15 min and viable GFP^{+} cells were sorted by FACS (Aria II; BD Biosciences) into DMEM containing 10% FBS.

Capturing of single cells and preparation of cDNA. Single embryonic lung epithelial cells were captured on a medium-sized (10–17 µm cell diameter) microfluidic RNA-seq or STA chip (Fluidigm) using the Fluidigm C1 system. To ensure unbiased and comprehensive profiling of all distal lung epithelial cells, an initial experiment was performed with a microfluidic chip with a 17–25-µm capture range; however, no cells with diameter greater than ~15 µm were captured, indicating that no major cell populations were missed by using the smaller capture range (Extended Data Fig. 2b). Cells were loaded onto the chip at a concentration of 300–500 cells µl^{-1}, stained for viability (LIVE/DEAD cell viability assay; Molecular Probes, Life Technologies) and imaged by phase-contrast and fluorescence microscopy to assess the number and viability of cells per capture site. Only single, live cells were included in the analysis. For RNaseq experiments, cDNAs were prepared on chip using the SMARTer Ultra Low RNA kit for Illumina (Clontech). ERCC (External RNA Controls Consortium) RNA spike-in Mix (Ambion, Life Technologies) was added to the lysis reaction and processed in parallel to cellular messenger RNA. For qPCR experiments, amplicons were prepared with pooled DELTAgene assays (Fluidigm) and Ambion (Life Technologies) Cells to CT lysis and pre-amplification kit, using the protocol provided by Fluidigm.

RNA-seq library construction. Single-cell cDNA size distribution and concentration was assessed on a capillary electrophoresis-based fragment analyser (Advanced Analytical). Illumina libraries were constructed in 96-well plates using the Illumina Nextera XT DNA Sample Preparation kit as described previously[7] using the protocol supplied by Fluidigm. For each C1 experiment, a bulk RNA control (about 200 cells) and a no-cell negative control were processed in parallel in PCR tubes, using the same reagent mixes as used on chip. Libraries were quantified by Agilent Bioanalyzer, using High Sensitivity DNA analysis kit, and also fluorometrically, using Qubit dsDNA HS Assay kits and a Qubit 2.0 Fluorometer (Invitrogen, Life Technologies).

DNA sequencing. Single-cell Nextera XT (Illumina) libraries of one experiment were pooled and sequenced 100 base pairs (bp) paired-end on Illumina HiSeq 2000 to a depth of $(2–6) \times 10^6$ reads (three replicate experiments of distal mouse lung epithelial cells at E18.5, one experiment at E14.5 and one experiment on adult AT2 cells) or 150 bp paired-end on Illumina MiSeq (one experiment at E16.5) to a depth of 100,000–550,000 reads with v3 chemistry. CASAVA 1.8.2 was used to separate out the data for each single cell by using unique barcode combinations from the Nextera XT preparation and to generate *.fastq files.

Microfluidic single-cell multiplexed qPCR. Single-cell multiplexed qPCR was performed in a M96 quantitative PCR DynamicArray on the Fluidigm Biomark instrument as described previously[26], using a panel of 96 DELTAgene assays (Fluidigm; Supplementary Table 2). In three of five single-cell qPCR experiments, ERCC spike-in transcripts (Ambion Live Technologies) were added to each single-cell lysis reaction on chip. Primer pairs for 6 of the 92 exogenous RNA spike-ins (ERCC spike-ins ERCC-00033, ERCC-00136, ERCC-00044, ERCC-00164, ERCC-00054 and ERCC-00074) were added to the preamplification reaction on chip and were subsequently used in the multiplexed qPCR experiment to detect the transcript level of each RNA spike-in. qPCR detection of the spike-in transcripts was later used to convert measured C_t values to approximate numbers of transcripts in a subset of 90 genes (Extended Data Fig. 7).

Processing, analysis and graphic display of single-cell RNA-seq data. Raw reads were pre-processed with the sequence-grooming tools FASTQC[30], cutadapt[31] and PRINSEQ[32] followed by sequence alignment with the Tuxedo suite (Bowtie[33], Bowtie2 (ref. 34), TopHat[25]) and SAMtools[35], using default settings. See Supplementary Data for information about the number of total reads and the percentage of mapped reads for each single cell. Transcript levels were quantified as fragments per kilobase of transcript per million mapped reads (FPKM) generated by TopHat/Cufflinks. Where depth matching was done, Seqtk (H. Li, https://github.com/lh3/seqtk/) was used to select raw reads randomly from each library, and the same pre-processing and alignment pipelines were used to obtain FPKM values for the depth-matched samples. Limit of detection of microfluidic single-cell RNA-seq was determined by analysing the correlation between the concentration of exogenous ERCC spike-in mRNA sequences and their respective mean FPKM values as measured by RNA-seq (Extended Data Fig. 3c). The spike-in sequences reflect a diverse range of sequence content and length, they have low homology with eukaryotic transcripts because they are from microbial sources, and they span a large range of concentrations to allow an empirical determination of the limit of detection[7,36,37]. The limit of detection was on the order of 0.5 molecules per reaction chamber, which is reflected as an FPKM value of ~1 (or 0 on a $\log_2$ scale). Transcripts with an FPKM value below or equal to 1 were therefore considered not expressed. Cells not expressing either of two housekeeping genes *Actb* and *Gapdh* (encoding β-actin and glyceraldehyde-3-phosphate dehydrogenase, respectively), or expressing them at less than three standard deviations below the mean, were scored as unhealthy and removed from the analysis. After applying this filter, a total of 80 cells remained for three replicate experiments at E18.5 (2 × pooled sibling lungs (20 and 26 cells), 1 × single lung (34 cells)), 45 cells remained for one experiment at E14.5, 27 cells remained for one experiment at E16.5 and 46 cells remained for an experiment of adult AT2 cells yielding 198 single cells in total.

For the lung epithelial cells at E18.5, we detected between 1,017 and 4,998 expressed genes per single cell, 10,946 in the union of all single cells and 8,653 in the 200 cell control bulk sample, indicating the heterogeneity of the analysed single cells. A total of 81 genes were commonly expressed in all single cells (Supplementary Table 1), which were mainly enriched for general processes such as translation.

FPKM values were converted to an approximate number of transcripts by using the correlation between the number of transcripts of exogenous spike-in mRNA sequences and their respective measured mean FPKM values (Extended Data Fig. 3c). The number of spike-in transcripts per single-cell lysis reaction was calculated from the concentration of each spike-in provided by the vendor (Ambion, Life Technologies), the approximate volume of the lysis chamber (10 nl) and the dilution of spike-in transcripts in the lysis reaction mix (×40,000). Transcript levels were converted to logarithmic space by taking $\log_2$(FPKM/number of transcripts). When calculating the characteristic single-cell expression (Extended Data Fig. 3b, d, f, g),

we used either the mean FPKM or the median FPKM value of each gene across all single cells transformed to the $\log_2$ space. To calculate the coefficient of variation of each gene across single cells (Extended Data Fig. 3b), the standard deviation of the $\log_2$-transformed FPKM values of a gene across all single cells was divided by the mean $\log_2$ FPKM value of the same gene.

Saturation plots (Extended Data Fig. 3a and 7a) were generated as described previously[7]. In brief, a corresponding number of millions of raw reads were randomly selected from each sample library and then, using the same alignment pipeline, FPKM values were called for each gene. This random subsampling was repeated, for each sample replicate, a total of four subsampled data sets per point, and the mean number of genes with an FPKM greater than 1 was plotted. For generating the 'single-cell ensemble' data set, raw reads from all the single-cell RNA-seq libraries were bioinformatically pooled to mimic a bulk RNA-seq experiment.

Custom R scripts[27] were used to perform principal component analysis (PCA), hierarchical clustering, Guilt-by-Association and permutation analysis as well as to construct violin plots, correlation plots and histograms. The scripts can be found in Supplementary Data. PCA analysis was performed on all cells using all genes expressed in more than two cells and with a variance in transcript level ($\log_2$(FPKM)) across all single cells greater than 0.5. Subsequently, genes with the highest principal-component loadings (highest absolute correlation coefficient with one of the first three or four principal components) were identified. Hierarchical clustering was performed on cells and on the genes identified by PCA, using Euclidean or correlation distance metric.

The specificity of the hierarchical clustering in Fig. 2a identifying five distinct cell populations was assessed with a permutation analysis approach. The sum of squares within groups (SSW) was therefore calculated for the cell grouping presented in Fig. 2a as well as for 50,000 random permutations thereof, keeping the size of cell groups and the total number of groups constant. With $x_{i,j}^k$ being the transcript level of gene j in cell i belonging to group k, the SSW can be calculated as

$$\mathrm{SSW} = \sum_{k=1}^{n} \sum_{j=1}^{m} (x_{i,j}^k - {}_i\overline{x}_j^k)^2$$

with ${}_i\overline{x}_j^k = \frac{1}{n}\sum_{i=1}^{n} x_{i,j}^k$ being the mean transcript level of gene j in all cells $i = 1, 2, \ldots, n$ belonging to group k. The SSWs for all 50,001 permutations were normally distributed and the SSW for our chosen clustering was significantly lower than that for all other permutations ($P = 2.89 \times 10^{-122}$).

When *Sftpc*$^+$ cells were isolated from all single-cell RNA-seq data sets (Fig. 4), a *Sftpc* transcript level of $\log_2$(FPKM) = 10 was chosen as threshold to separate cells with background *Sftpc* expression from cells with high *Sftpc* expression (referred to as *Sftpc*$^+$ cells).

To search for further previously unknown cell-type markers and cell-type specific transcription factors or receptors/ligands beside the genes identified by PCA, we defined a 'perfect marker gene' for each cell type with a high transcript level ($\log_2$(FPKM) = 10) in all cells of the respective cell type, and with no expression (FPKM = 0) in all other cells. We then determined the pairwise Pearson correlation between the single-cell expression profile of each perfect marker gene and every other transcribed gene. The list of murine transcription factors that were screened for cell-type specificity was obtained from the online animal transcription factor database (http://www.bioguo.org/AnimalTFDB/) (ref. 38). All genes identified as cell-type-specific by PCA analysis and hierarchical clustering (see above) also had a high Pearson correlation coefficient with the corresponding perfect marker gene. The Pearson correlation coefficients for the most strongly correlating genes are shown in Supplementary Data together with information about the top 30 genes per cell type regarding previous detection in cell types in the lung and regarding available literature or known mouse knockout phenotypes.

Guilt-by-Association analysis[39] was used to calculate the probability of observing a given co-expression of two genes by chance. Gene expression values were therefore scaled gene-by-gene by mean-centring and dividing by the standard deviation of the respective gene across all single cells, and a binary expression matrix was constructed by defining a gene as expressed in a given cell if the scaled expression level was greater than or equal to 0, and as not expressed if it was smaller than 0. Pairwise comparisons were performed between the perfect marker gene for each of the four mature cell types (AT1, AT2, Clara and ciliated) and all other genes expressed in at least one cell (10,946 genes in total). P values were calculated with the hypergeometric distribution as described in ref. 37, and multiple testing was accounted for by the Benjamini–Hochberg method (Fig. 2b and Supplementary Data).

Gene ontology and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were performed with DAVID informatics resources 6.7 of the National Institute of Allergy and Infectious Diseases of the NIH[21,40] (Supplementary Data).

Analysis and graphic display of microfluidic single-cell multiplexed qPCR data. Single-cell multiplexed qPCR data were analysed and displayed by using custom R scripts[27]. qPCR experiments were performed for E16.5 (two biological replicates), E18.5 (two biological replicates) distal lung epithelial cells and for adult AT2 cells (one replicate). The limit of detection of multiplexed qPCR values was determined as 22 threshold cycles (C_t) by a calibration experiment with 16-fold serial dilutions of total lung cDNA and six replicates for each concentration. Genes that were not expressed were given a value higher than the limit of detection and the limit of detection was subtracted from all C_t values to transform C_t values to $\log_2$ expression values ($\log_2\mathrm{Ex} = C_{t,\mathrm{LoD}} - C_t$, $C_{t,\mathrm{LoD}} = 22$). Cells not expressing either of two housekeeping genes *Actb* and *Gapdh*, or expressing them at less than three standard deviations below the mean, were scored as unhealthy and removed from the analysis. After applying this filter, 74 single cells remained for two experiments at E18.5, 33 cells for two experiments at E16.5, and 48 cells for the experiment with adult AT2 cells. In all experiments, cells were isolated from pooled lungs from one litter (five to nine lungs). To combine experiments from different chips for the same embryonic time point, the expression value of each gene for a given cell was normalized to the median gene expression value of that cell. Normalized gene expression values were further scaled gene by gene by mean-centring and dividing by the standard deviation of expressing cells. PCA and hierarchical clustering using Euclidean distance metric were performed in R for all cells, using ten canonical marker genes for bronchiolar and alveolar cells (*Abca3*, *Sftpb*, *Muc1*, *Sftpc*, *Lyz2*, *Aqp5*, *Pdpn*, *Ager*, *Foxj1* and *Scgb1a1*).

Immunofluorescence. E18.5 lungs were removed *en bloc*; for whole-mount staining, the tip of the accessory lobe was excised. Lungs and tips were immersion-fixed overnight in paraformaldehyde (4% in PBS) at 4 °C, and then dehydrated and stored in methanol at −20 °C until being stained. Lungs of adult mice were collected as above except that after clearance of the pulmonary vasculature, the ventral trachea was incised and cannulated, and lungs were gently inflated to full capacity with molten low-melting-point agarose (2% in PBS; Sigma). Ice-cold PBS was dripped into the thorax to solidify the agarose, and inflated lungs were removed *en bloc* and processed as above. E18.5 lungs were rehydrated, cryoprotected overnight in 30% sucrose at 4 °C, submerged in OCT (Tissue Tek) in an embedding mould, frozen on solid CO_2, then stored at −80 °C. Sections of thickness 10 μm obtained with a cryostat (Leica CM3050S) were collected on chambered glass slides and stored at 4 °C before being stained.

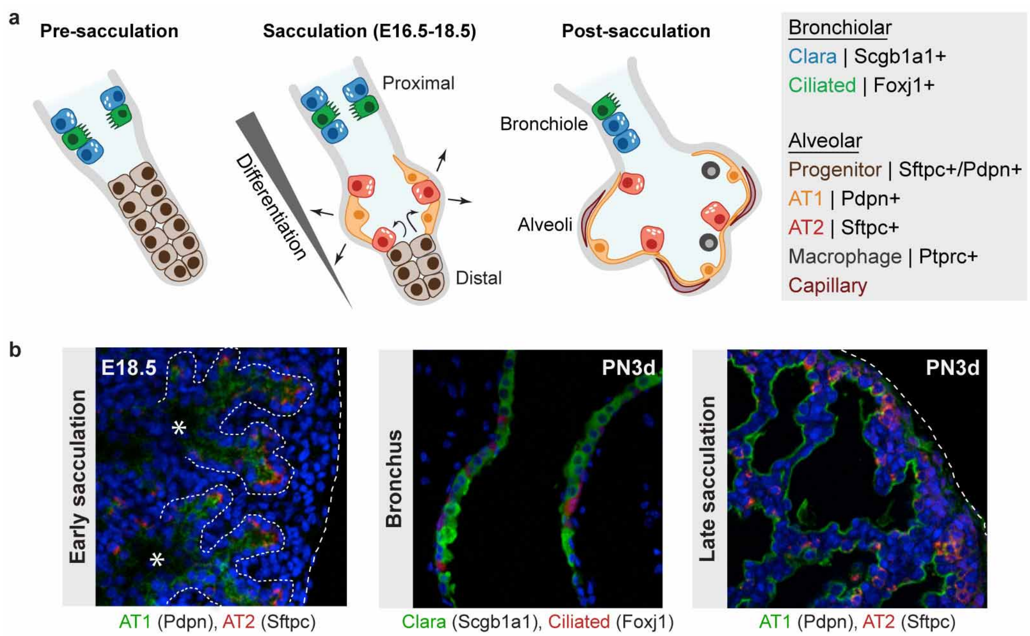
Similar immunofluorescence protocols were used on whole mounts as on sections, except that incubation times were increased to compensate for tissue thickness. Lung tissue was permeabilized (10 min, PBS containing 0.3% Triton X-100), washed (three 5-min washes in PBS containing 0.1% Tween 20) and blocked (1 h, PBS containing 10% donkey serum) before incubation overnight with primary antibody. Adult lungs did not require further permeabilization. Primary antibodies against the following antigens (used at 1:200 dilution unless otherwise noted) were pro-SftpC (rabbit; Chemicon AB3786), Pdpn (hamster; DSHB 8.1.1), E-cadherin (rat; Zymed ECCD-2), Rage (rat; R&D), Scgb1a1 (rabbit; Upstate), Lamp3 (sheep, AF4584; R&D), S100a6 (rat, DDX0192; Dendritics) and Krt15 (mouse, LHK15; SCBT) directly conjugated to a fluorophore in accordance with the manufacturer's instructions (Alexa Fluor Antibody Labelling Kit; Invitrogen). After further washing, sections were incubated with appropriate secondary antibodies conjugated to an Alexa fluorophore (donkey A488, A555 or A633; Invitrogen) as well as DAPI (5 ng ml^{-1}) for 1 h, followed by washing and mounting in Vectashield (Vector). Lung tissues were imaged with a laser scanning confocal microscope (LSM 780; Zeiss).

***In situ* hybridization.** For *in situ* hybridizations, embryonic lungs were collected as for immunostaining (see above), washed briefly in PBS (autoclaved, diethyl pyrocarbonate-treated) and snap-frozen in OCT for sectioning. Sections of thickness 10 μm were generated on the cryostat and stored at −20 °C before further processing. To validate AT2-specific expression of *Egfl6* RNA by *in situ* hybridization, sections were transported on dry ice to a company specializing in processing and imaging dual *in situ* hybridized samples, in accordance with the company's reported protocol (ACD's RNAscope *In situ* Hybridization Technology). To validate the expression of *Sftpc* and explore its spatial expression pattern in the embryonic mouse lung (E11.5, E13.5 and E14.5), *in situ* hybridizations were performed on whole-mount mouse lungs as described previously[41].

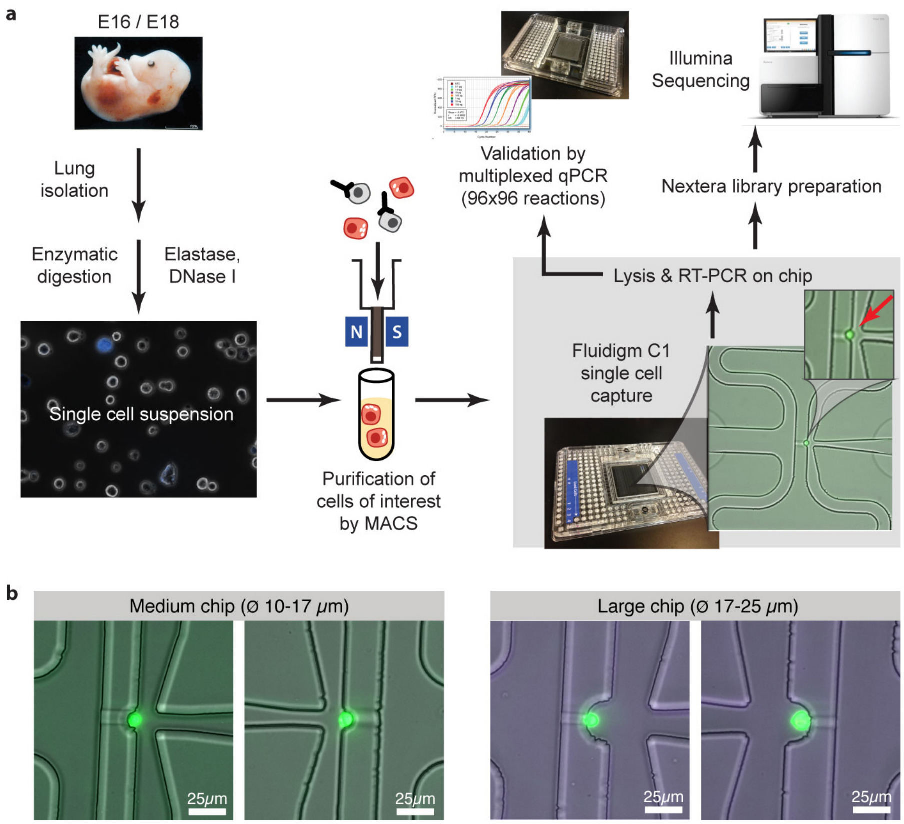
Validation of *Hopx* as AT1 marker gene by transgenic labelling. Cells actively transcribing Hopx in the adult mouse lung were labelled by injecting 2 mg of tamoxifen (Sigma) in corn oil at 20 mg ml^{-1} concentration intraperitoneally into postnatal day-28 Hopx-Cre-ERT2$^{+/-}$ mTmG$^{+/-}$ mice. Three days later, lungs were collected as described above, fixed overnight in paraformaldehyde (4% in PBS) at 4 °C and stored in 80% glycerol at 4 °C before imaging with a laser scanning confocal microscope (LSM 780; Zeiss) with a 0.8 numerical aperture, 25× oil-immersion objective and confocal z-sections with a thickness of 2.3 μm.

28. Chapman, H. A. *et al.* Integrin α6β4 identifies an adult distal lung epithelial population with regenerative potential in mice. *J. Clin. Invest.* **121,** 2855–2862 (2011).

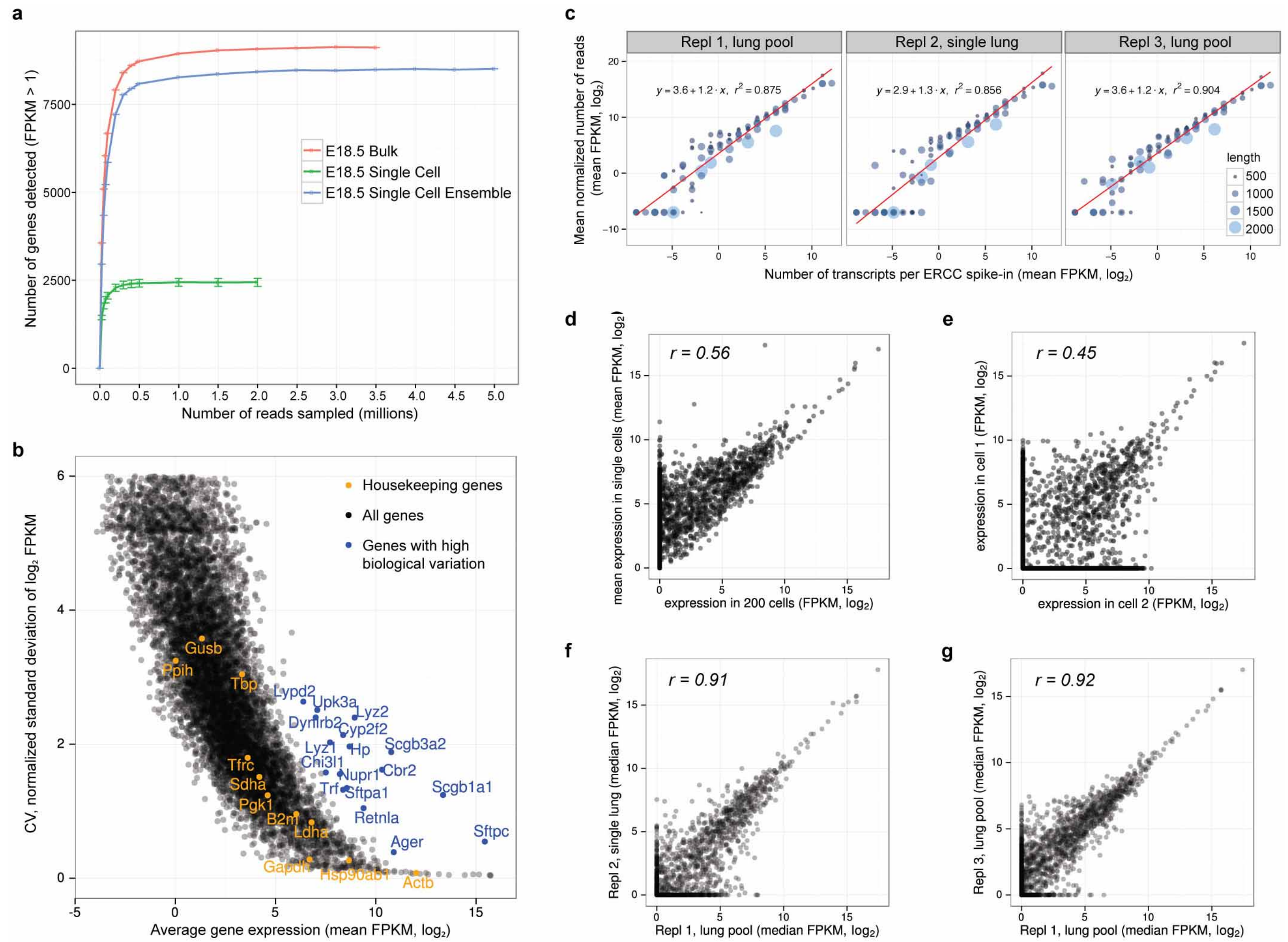
29. Takeda, N. *et al.* Interconversion between intestinal stem cell populations in distinct niches. *Science* **334,** 1420–1424 (2011).
30. Babraham Institute. Babraham Bioinformatics. FASTQC. http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc.
31. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17,** 10–12 (2011).
32. Schmieder, R. & Edwards, R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27,** 863–864 (2011).
33. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10,** R25 (2009).
34. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9,** 357–359 (2012).
35. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).
36. Baker, S. C. *et al.* The External RNA Controls Consortium: a progress report. *Nature Methods* **2,** 731–734 (2005).
37. Jiang, L. *et al.* Synthetic spike-in standards for RNA-seq experiments. *Genome Res.* **21,** 1543–1551 (2011).
38. Zhang, H.-M. *et al.* AnimalTFDB: a comprehensive animal transcription factor database. *Nucleic Acids Res.* **40,** D144–D149 (2012).
39. Walker, M. G., Volkmuth, W., Sprinzak, E., Hodgson, D. & Klingler, T. Prediction of gene function by genome-scale expression analysis: prostate cancer-associated genes. *Genome Res.* **9,** 1198–1203 (1999).
40. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37,** 1–13 (2009).
41. Greif, D. M. *et al.* Radial construction of an arterial wall. *Dev. Cell* **23,** 482–493 (2012).



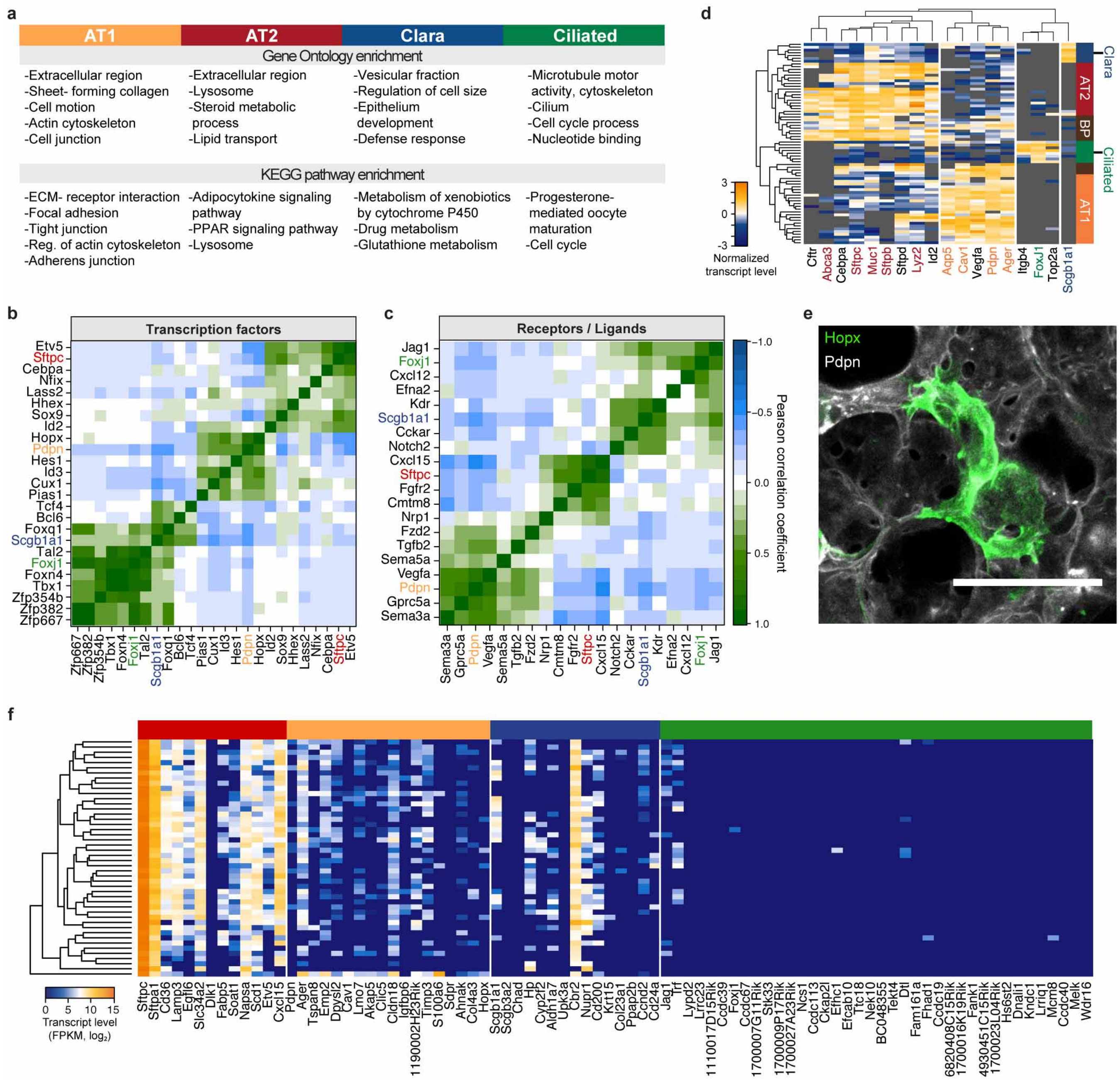
Extended Data Figure 1 | Schematic illustration of the process of sacculation. **a**, Schematic illustration of morphological and molecular changes of the distal airways during development. Cell differentiation progresses in a directional manner from the bronchio-alveolar junction (proximal) to the distal tip (distal) of each terminal airway; progenitor cells therefore persist the longest at the tips. Ciliated (green) and Clara (blue) cells mature first, followed by the differentiation of flat alveolar type 1 (AT1, orange) and cuboidal type 2 (AT2, red) cells from cuboidal alveolar progenitors during sacculation (E16–18.5), when distal airway tubules widen as nascent AT1 cells flatten to form the gas-exchange surface. **b**, Micrographs of alveolar (E18.5, postnatal 3 days (PN3d)) and bronchiolar (PN3d) sections of a mouse lung co-stained for Clara (Scgb1a1, green) and ciliated (Foxj1, red) cell markers as well as AT1 (Pdpn, green) and AT2 (Sftpc, red) specific markers. Progenitor cells at the tips of sacculating alveoli are detected by an overlap of AT1 and AT2 specific markers. Newly forming alveolar sacs are marked by asterisks.



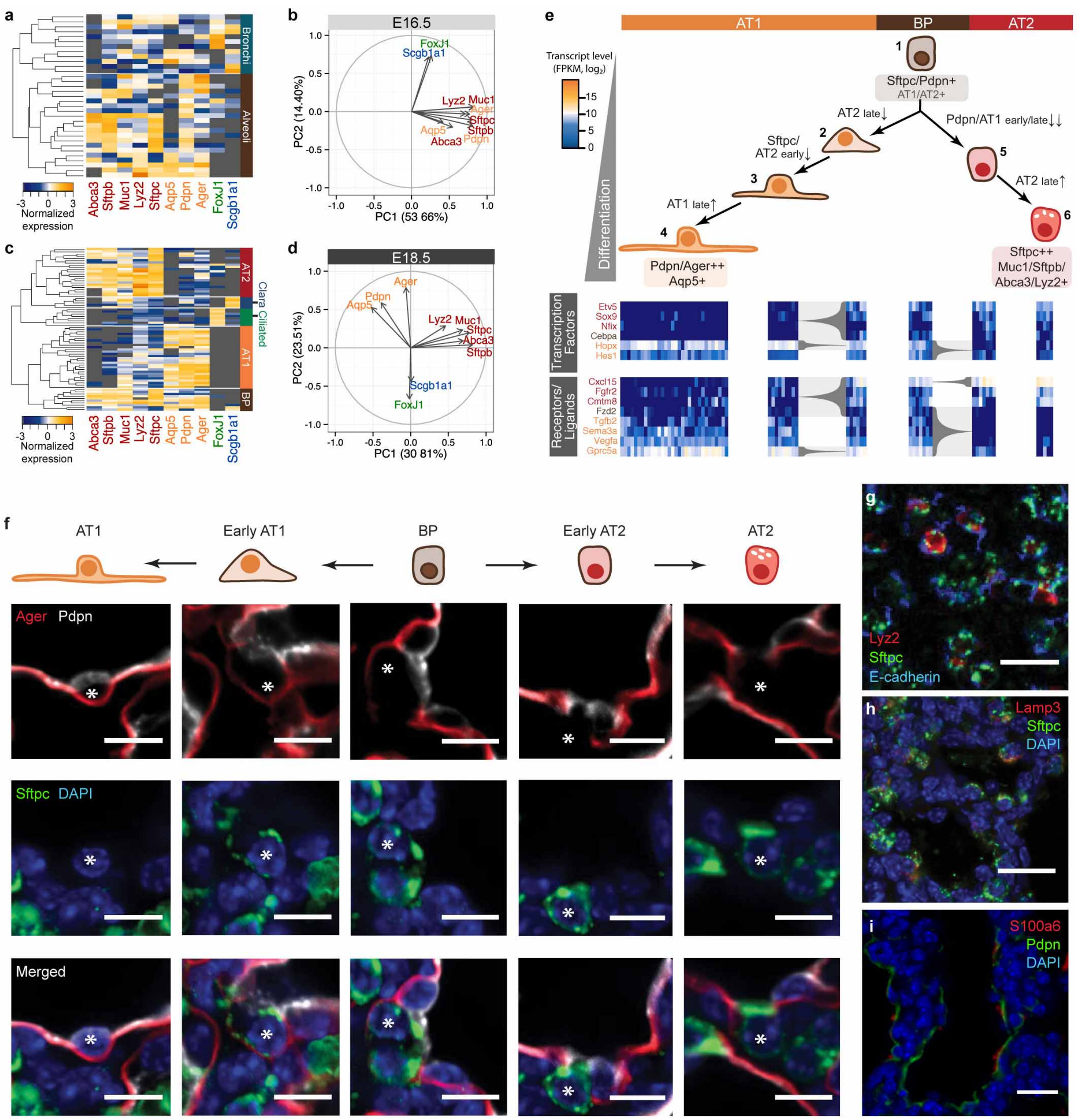
Extended Data Figure 2 | Single-cell transcriptomics analysis workflow. **a**, Workflow of single-cell transcriptomics analysis of mouse lung epithelial cells. A single captured lung epithelial cell stained with Alexa488 for EpCAM (green) is indicated by a red arrow. **b**, Single lung epithelial cells captured in microfluidic chips with capture sites designed to trap cells with a diameter of 10–17 μm (medium, left) or 17–25 μm (large, right). Cells were stained for viability with Calcein AM. Even cells captured by the large chip did not exceed a diameter of ~15 μm, indicating that the medium-sized chips are sufficient for comprehensively profiling distal mouse lung epithelial cells.



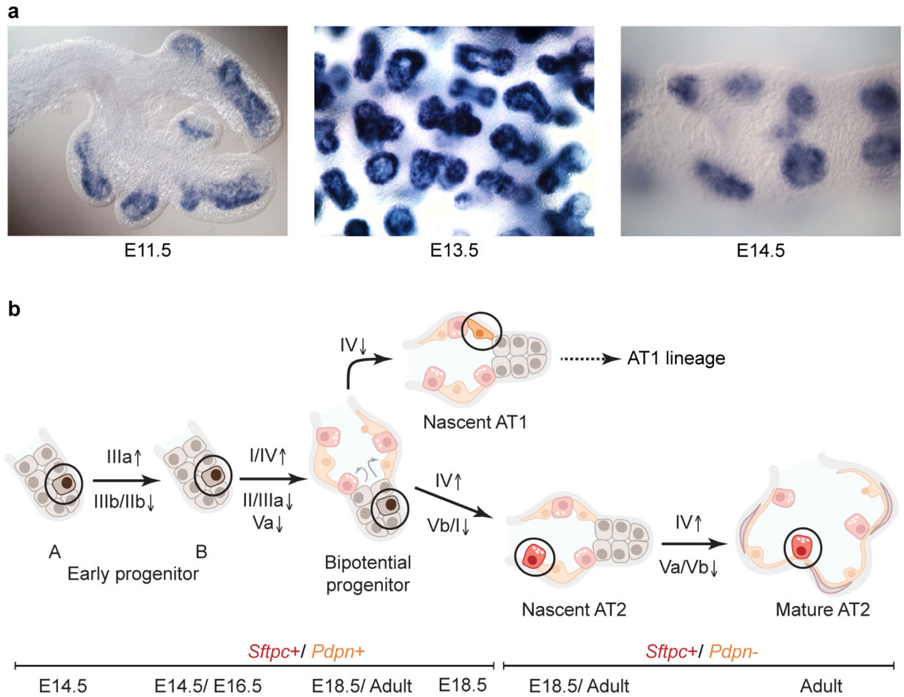
Extended Data Figure 3 | Assessment of required sequencing depth, technical and biological variation, dynamic range and reproducibility of single-cell RNA-seq data of 80 single distal lung epithelial cells at E18.5. **a**, Saturation analysis reveals the sequencing depth required for the detection of most genes expressed by single cells. To detect most expressed genes, single-cell RNA-seq libraries have to be sequenced only to a depth of about 10^6 reads, whereas libraries of bulk samples have to be sequenced more deeply. The number of genes detected in the ensemble of all single cells (synthetic bulk) is comparable to the number of genes detected in the true bulk experiment. Each point on the saturation curve was generated by randomly selecting a number of raw reads from each sample library (bulk, 200 cell bulk library; single cell, single-cell RNA-seq libraries of 80 lung epithelial cells; single-cell ensemble, bioinformatically pooled single-cell libraries) and then using the same alignment pipeline to call genes with a mean FPKM of more than 1. Each point represents four replicate subsamplings; error bars represent s.e.m. **b**, Technical noise and biological variation in single-cell RNA-seq data. Relationship between mean expression level and coefficient of variation for 10,946 genes in single embryonic lung epithelial cells. Several genes show strong biological variation (blue): they show higher variability than the average noise at a given average gene expression. Housekeeping genes are shown in yellow. **c**, Average detected transcript levels (mean FPKM, $\log_2$) for 92 ERCC RNA spike-ins as a function of provided number of molecules per lysis reaction for each of the three independent single-cell RNA-seq experiments performed at E18.5. Linear regression fits through data points are shown. The length of each ERCC RNA spike-in transcript is encoded in the size and colour of the data points. No particular bias towards the detection of shorter versus longer transcripts is observed. The method shows single transcript sensitivity as well as a dynamic range of approximately six orders of magnitude, in agreement with a previous study evaluating microfluidic single-cell RNA-seq[7]. **d**, **e**, Correlation between transcript levels of a 200-cell population and median transcript levels of single cells of the same pool of embryonic lungs (**d**), and transcript levels of two single AT2 cells (**e**). *r*, Pearson correlation coefficients. **f**, **g**, Correlation between transcript levels of all genes detected in the single lung and the pooled lung experiment (**f**) and between transcript levels of all genes detected in the two independent experiments on pooled embryonic lungs (**g**). Pearson correlation coefficients *r* are given.



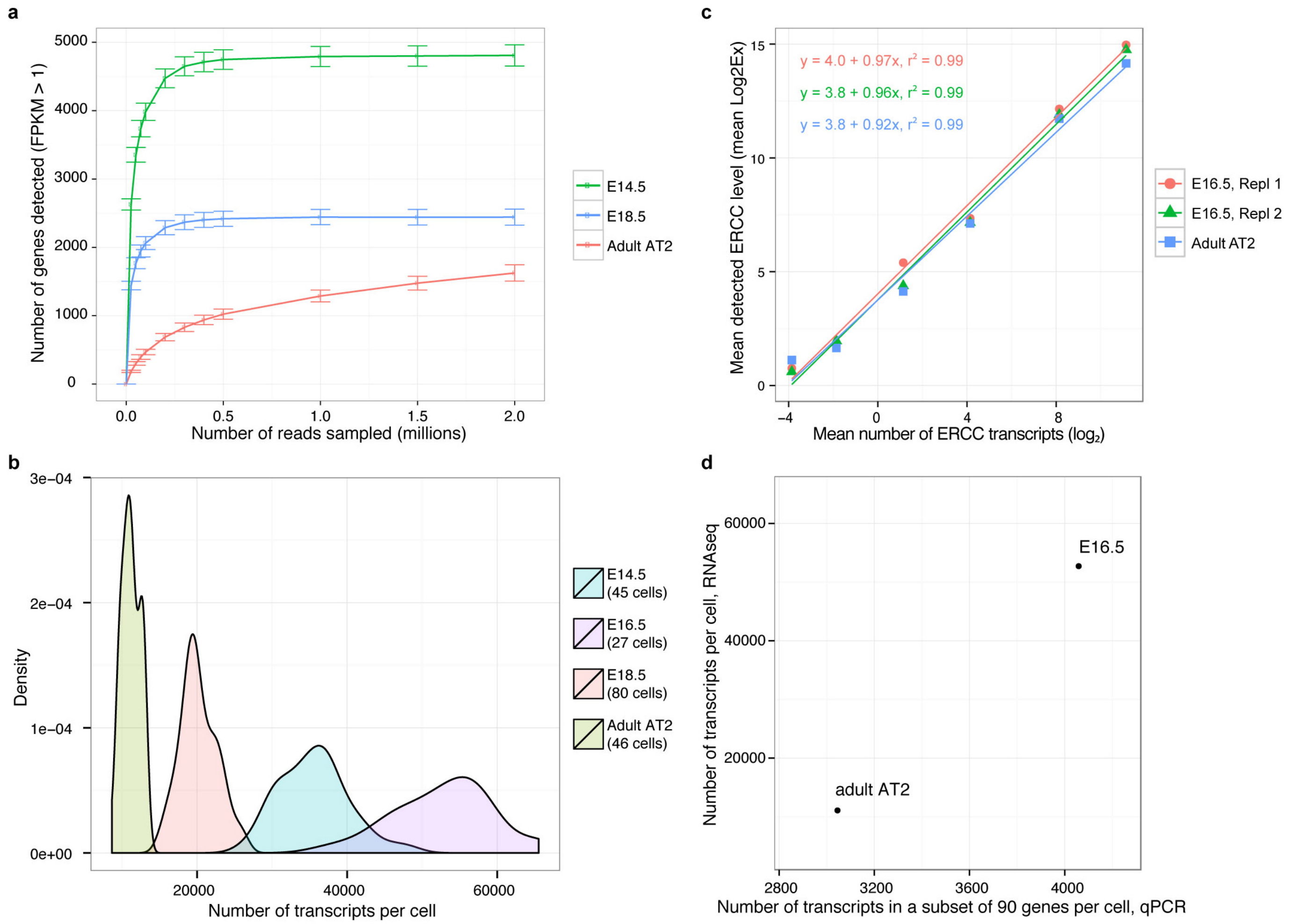
Extended Data Figure 4 | Lineage-specific genes identified by single-cell transcriptome analysis allow functional description of individual distal lung epithelial cell populations. **a**, Results of gene ontology (GO) and KEGG pathway enrichment analyses for distal lung epithelial cell types based on lineage-specific genes identified by single-cell RNA-seq of 80 E18.5 distal lung epithelial cells (Supplementary Data). **b**, **c**, Correlograms visualizing correlation of single-cell gene expression profiles between transcription factors (**b**) or receptors/ligands (**c**) and the major canonical marker genes for bronchiolar and alveolar lineages (AT1: *Pdpn*; AT2: *Sftpc*; Clara: *Scgb1a1*; ciliated: *Foxj1*). The colour bar denotes the Pearson correlation coefficient from −1 (blue, anticorrelated genes) to 1 (green, positively correlated genes). **d**, Validation of previously unknown marker genes by single-cell multiplexed qPCR on 74 single cells isolated from the distal mouse lung epithelium at E18.5. Lineage-specific expression of seven new marker genes is shown by clustering with known markers for respective lineages (AT2, red, previously unknown: *Cftr*, *Cebpa*, *Sftpd* and *Id2*; AT1, orange, previously unknown: *Vegfa*; ciliated, green, previously unknown: *Itgb4* and *Top2a*; Clara, blue). **e**, Validation of *Hopx* expression in AT1 cells. A lung section from a transgenic *Hopx*>*GFP* adult mouse (*Hopx-Cre-ERT2*$^{+/-}$;*mTmG*$^{+/tg}$) was co-stained for AT1 marker Pdpn. Maximum-intensity projections of confocal *z* stacks show that AT1 cells expressing the membrane-localized GFP reporter (green) also express Pdpn (white). Scale bar, 50 µm. **f**, Hierarchical clustering of 46 transgenically labelled mature *Sftpc*$^{+}$ AT2 cells, isolated by FACS from adult mouse lung. Most genes identified as AT2 lineage-specific from single-cell transcriptomes at E18.5 are transcribed also by mature AT2 cells. In contrast, no or low expression is observed in mature AT2 cells for the genes specific to the other alveolar or bronchiolar lineages as identified from single-cell RNA-seq data at E18.5.



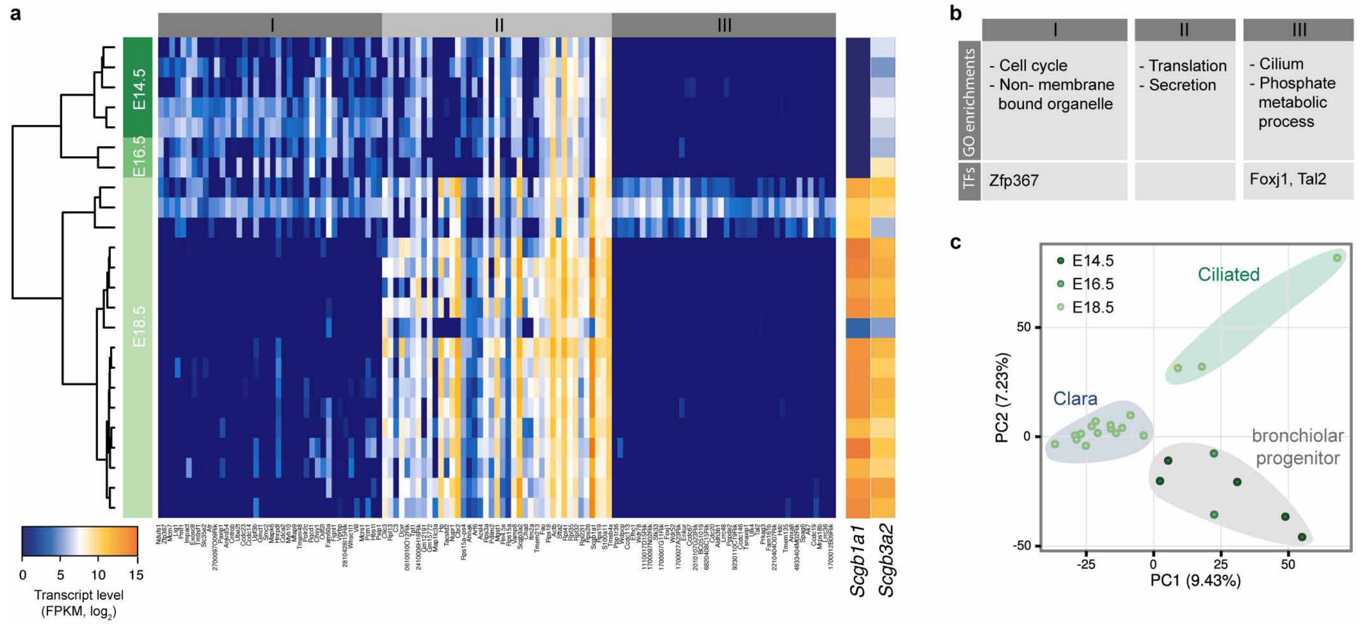
Extended Data Figure 5 | Molecular profiles distinguish developmental intermediates during the differentiation of AT1 and AT2 cells from a common BP. **a**, Hierarchical clustering of multiplexed qPCR gene expression data for 33 single cells from E16.5 lung epithelium (CD45^{-}/EpCAM^{+}) suggests the presence at this time point of two major cell lineages, bronchiolar (cyan) and alveolar (brown) progenitors. Note that alveolar progenitors express a subset of both AT1 and AT2 marker genes. **b**, PCA of multiplexed qPCR data of lung epithelial cells at E16.5 identifies two gene groups in contrast to three observed at E18.5 (Fig. 1c). AT1 and AT2 specific marker genes do not segregate into distinct populations at E16.5. **c**, Hierarchical clustering of multiplexed qPCR gene expression data for 74 single embryonic lung epithelial cells (CD45^{-}/EpCAM^{+}) at E18.5 shows multiple distinct cell populations consistent with RNA-sequencing data at this time point: BP, AT1, AT2, Clara and ciliated cells. Each row represents a single cell and each column a gene. Cells are clustered on the basis of expression of marker genes for alveolar and bronchiolar lineages (AT2: *Abca3*, *Sftpb*, *Muc1*, *Lyz2*, *Sftpc*; AT1: *Aqp5*, *Pdpn*, *Ager*; ciliated: *Foxj1*; Clara: *Scgb1a1*). **d**, PCA of multiplexed qPCR data replicates gene families found by single-cell RNA-seq at E18.5. Gene groups were characterized on the basis of differential correlation with the first two principal components. **e**, Developmental sequence of AT1 (orange) and AT2 (red) specification from a common BP (brown). Two and three maturation intermediates were identified in the specification process of AT2 and AT1 cell types, respectively, on the basis of the expression of known and previously unknown marker genes for both alveolar lineages measured by single-cell RNA-seq (Fig. 3). Transcription factors and receptors/ligands shown here were found to be expressed in BP cells and subsequently restricted to one of the alveolar lineages. Arrows, differentiation pathway; grey braces, change in transcript level of respective genes with tip pointing towards lower expression. **f–i**, Protein level heterogeneity of alveolar epithelial markers during sacculation. **f**, Immunofluorescent micrograph from an E19.5 lung with mature AT1 and AT2 cells stained for their respective markers (Pdpn (white) and Ager (red) for AT1; Sftpc (green) for AT2). BPs are positive for all three markers. Cells in intermediate states are observed, such as early AT1 (Pdpn and Ager positive, Sftpc low) and early AT2 cells (Sftpc positive, and either Pdpn positve/Ager low or Pdpn low/Ager negative). Scale bar, 10 µm. **g**, Markers of late AT2 cells are expressed heterogeneously at E18.5. Immunofluorescence micrograph of a lung from a Lyz2–enhanced green fluorescent protein (eGFP) transgenic mouse, in which within the epithelium (E-cadherin, blue) only a subset of Sftpc (green)-positive AT2 cells are Lyz2 (red)-positive. Scale bar, 20 µm. **h**, Immunofluorescent staining of E18.5 lung tissue for Lamp3 (red) shows heterogeneous expression of Lamp3 in Sftpc-positive cells (green): Proximal cells show higher Lamp3 expression than distal cells. Blue, DAPI-stained nuclei. Scale bar, 20 µm. **i**, Immunofluorescent staining of E18.5 lung tissue for S100a6 (red) shows heterogeneous expression of the secreted protein S100a6 in Pdpn-positve cells (green). Blue, DAPI-stained nuclei. Scale bar, 20 µm.



Extended Data Figure 6 | Following *Sftpc*-expressing cells throughout their life cycle. **a**, Whole-mount *in situ* hybridizations of embryonic mouse lungs at E11.5, E13.5 and E14.5 using probes against *Sftpc* mRNA show expression of *Sftpc* specific to the tips of the epithelial tree branches. Moreover, variations in signal intensity indicate heterogeneity in the level of *Sftpc* expression across cells, which is in agreement with our single-cell RNA-seq data of *Sftpc*$^{+}$ cells at E14.5 (see Fig. 4a). **b**, Diagram of the different transcriptional states in the specification of an AT2 cell as identified by single-cell RNA-seq of *Sftpc*$^{+}$ cells from distal mouse lung epithelium of embryonic (E14.5, E16.5 and E18.5) and adult mice. The cell undergoes a transition from an early (A) and late (B) early progenitor state into a BP state before either taking the AT1 fate (nascent AT1), or following the AT2 pathway to become a nascent and finally a mature AT2 cell. Groups of genes turning on/up or off/down during the individual transitions are shown above and below each arrow, respectively (Fig. 4a and Supplementary Data). Whereas EP and BP cells are double positive for *Sftpc* and *Pdpn*, nascent and mature AT2 cells express *Sftpc* but turn off expression of the AT1 marker *Pdpn*. The developmental time points at which the individual cell states were detected, and their putative locations, are shown.



Extended Data Figure 7 | The number of unique genes and the total number of transcripts expressed by a single cell strongly correlates with its differentiation state. **a**, Saturation analysis of single-cell RNA-seq data of lung epithelial cells at different embryonic and adult time points (E14.5, E18.5 and adult AT2) reveals that the number of unique genes expressed by single lung epithelial cells decreases with progressing differentiation state. Distal lung epithelial cells at E14.5 express more than 6,000 genes, whereas cells at E18.5 express about 3,000 genes, and mature AT2 cells only about 2,000 genes. Each point on the saturation curve was generated by randomly selecting a number of raw reads from each sample library and then using the same alignment pipeline to call genes with a mean FPKM of more than 1. Each point represents four replicate subsamplings. Error bars represent s.e.m. All libraries were sequenced to a depth of at least 2×10^6 reads. **b**, Single-cell RNA-seq reveals that the total number of transcripts expressed by single cells decreases with increasing differentiation state of the cell. The number of transcripts per cell was calculated from the FPKM values of all genes in each cell, using the correlation between number of transcripts of exogenous spike-in mRNA sequences and their respective measured mean FPKM values (example calibration curves are shown in Extended Data Fig. 3c for three replicates at E18.5). Area-normalized density distributions are shown for embryonic cells at E14.5 (45 cells), E16.5 (27 cells) and E18.5 (80 cells), and for 46 *Sftpc*$^+$ adult AT2 cells. The number of transcripts is highest in lung epithelial progenitor cells at E16.5 and E14.5 and decreases in cells at E18.5 and even further in mature AT2 cells. Note that single-cell RNA-seq libraries for E14.5, E18.5 and adult AT2 cells were sequenced to a depth of $(2\text{–}6) \times 10^6$ reads, whereas the libraries for cells at E16.5 were sequenced to a lower depth of 100,000–550,000 reads. **c**, Calibration of C_t values measured by single-cell qPCR to number of molecules. Average detected transcript levels ($\log_2\text{Ex} = C_{t,\text{LoD}} - C_t$, $C_{t,\text{LoD}} = 22$) for six ERCC RNA spike-ins as a function of provided number of molecules per lysis reaction for each of three independent single-cell qPCR experiments performed on embryonic (E16.5, two replicates; red and green) and adult mouse lung (adult AT2, one replicate; blue). Linear regression fits through data points and corresponding equations are shown and were used to convert C_t values measured by qPCR into numbers of transcripts. **d**, Single-cell qPCR confirms the presence of a higher number of transcripts in lung epithelial progenitor cells in comparison with fully differentiated alveolar epithelial cells. The median number of transcripts per cell as detected by single-cell RNA-seq (*y* axis) and by single-cell multiplexed qPCR of 90 genes (*x* axis) is shown for distal lung epithelial cells at E16.5 (qPCR, 33 cells; RNA-seq, 27 cells) and mature AT2 cells (qPCR, 48 cells; RNA-seq, 46 cells).



Extended Data Figure 8 | Transcriptional states during the early lifetime of the Clara cell lineage identified by single-cell RNA-seq of *Scgb3a2*$^+$ cells at E14.5, E16.5 and E18.5. **a**, Hierarchical clustering of 24 *Scgb3a2*-positive cells from distal mouse lung epithelium at different embryonic time points (E14.5, E16.5 and E18.5) based on the genes with highest principal-component loadings in an unbiased PCA analysis of all cells and all genes (shown in **c**). Cells are shown in rows, genes in columns. Cells cluster into three major groups. *Scgb3a2* and *Scgb1a1* transcript levels are shown in bars on the right. Whereas canonical Clara cell marker *Scgb1a1* is first detected at E18.5, *Scgb3a2* is detected as early as E14.5, suggesting that it is an early Clara cell marker. **b**, Gene Ontology (GO) enrichments of the three different gene clusters as well as transcription factors (TFs) belonging to the different groups of genes. **c**, PCA analysis of all *Scgb3a2*-positive cells and all genes identifies three different cell populations that were identified as bronchiolar progenitor as well as Clara and ciliated cells.

doi:10.1038/nature13084

Disruption of Mediator rescues the stunted growth of a lignin-deficient *Arabidopsis* mutant

Nicholas D. Bonawitz[1]†, Jeong Im Kim[1], Yuki Tobimatsu[2], Peter N. Ciesielski[3], Nickolas A. Anderson[1], Eduardo Ximenes[4], Junko Maeda[1]†, John Ralph[2,5,6], Bryon S. Donohoe[3], Michael Ladisch[4,7] & Clint Chapple[1]

Lignin is a phenylpropanoid-derived heteropolymer important for the strength and rigidity of the plant secondary cell wall[1,2]. Genetic disruption of lignin biosynthesis has been proposed as a means to improve forage and bioenergy crops, but frequently results in stunted growth and developmental abnormalities, the mechanisms of which are poorly understood[3]. Here we show that the phenotype of a lignin-deficient *Arabidopsis* mutant is dependent on the transcriptional co-regulatory complex, Mediator. Disruption of the Mediator complex subunits MED5a (also known as REF4) and MED5b (also known as RFR1) rescues the stunted growth, lignin deficiency and widespread changes in gene expression seen in the phenylpropanoid pathway mutant *ref8*, without restoring the synthesis of guaiacyl and syringyl lignin subunits. Cell walls of rescued *med5a/5b ref8* plants instead contain a novel lignin consisting almost exclusively of *p*-hydroxyphenyl lignin subunits, and moreover exhibit substantially facilitated polysaccharide saccharification. These results demonstrate that guaiacyl and syringyl lignin subunits are largely dispensable for normal growth and development, implicate Mediator in an active transcriptional process responsible for dwarfing and inhibition of lignin biosynthesis, and suggest that the transcription machinery and signalling pathways responding to cell wall defects may be important targets to include in efforts to reduce biomass recalcitrance.

The phenolic polymer lignin has a central role in both the fitness of plants and their utility to humans. During development, lignin is deposited along with cellulose and hemicelluloses in the secondary cell walls of structural fibres and water-conducting cells, where it is essential for their strength and rigidity[1,2]. Lignin deposition is also induced at sites of wounding or pathogen attack[4], impeding the action of fungal and bacterial cellulolytic enzymes and thus inhibiting pathogen invasion of surrounding tissues. Lignin limits the value of biomass crops used for forage or lignocellulosic biofuel production by interfering with the breakdown of cellulose and other wall polysaccharides to simple sugars[5]. Attempts to reduce biomass recalcitrance through genetic manipulation of lignin deposition have met with some success, but the stunted growth of many of the resulting plants and the associated yield penalty have made the use of similar genetic modifications in commercial biomass crops problematic[3]. The molecular events underlying the dwarfing of lignin biosynthetic mutants are at present poorly understood, but may involve lignin deficiency *per se*, loss (or gain) of a related non-lignin metabolite, or the activation of a cell wall integrity monitoring pathway analogous to that triggered by the receptor-like tyrosine kinase THE1 in response to cellulose deficiency[6]. Understanding and potentially mitigating the pleiotropic phenotypes of lignin biosynthetic mutants could markedly alter the economic feasibility of lignocellulosic biofuel production.

The lignin polymer is derived mainly from three chemically related but distinct hydroxycinnamyl alcohols that, after oxidation and polymerization in the secondary cell wall, give rise to *p*-hydroxyphenyl (H), guaiacyl (G) and syringyl (S) lignin subunits[7] (Fig. 1a). Although S lignin is dispensable for normal growth[8], disruption of shared phenylpropanoid pathway enzymes upstream of both G and S lignin leads to considerable developmental abnormalities[9–13]. For instance, the *Arabidopsis* mutant *reduced epidermal fluorescence 8-1* (*ref8-1*) is severely dwarfed and sterile due to a missense mutation in the gene encoding *p*-coumaroylshikimate 3′-hydroxylase (C3′H) that reduces, but does not eliminate its activity[9,10,14].

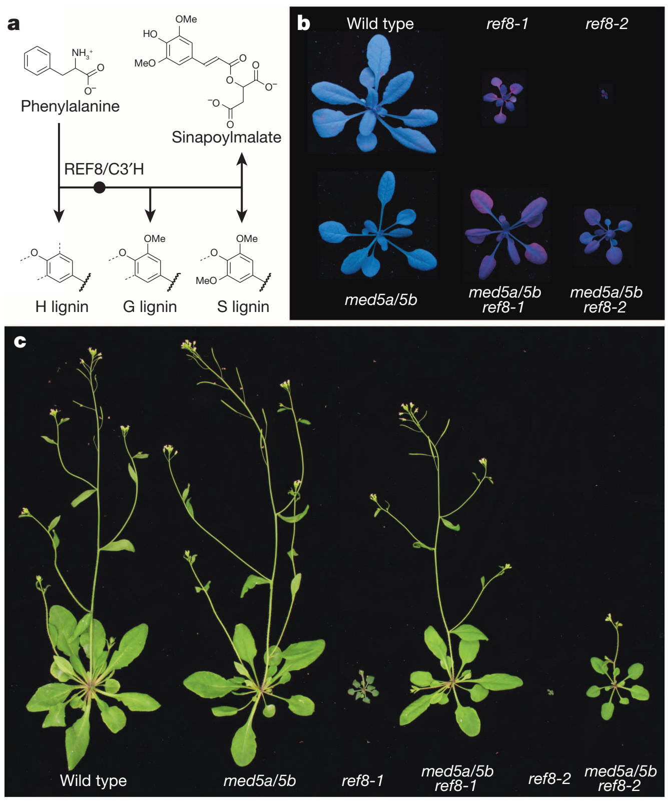


Figure 1 | Disruption of *MED5a/5b* rescues growth and fertility of *Arabidopsis ref8* mutants. a, Simplified view of the phenylpropanoid pathway, showing the position of REF8/C3′H in relation to the pathway branches leading to H, G and S lignin subunits and sinapoylmalate. **b**, Three-week-old wild-type, *ref8-1*, *ref8-2*, *med5a/5b*, *med5a/5b ref8-1* and *med5a/5b ref8-2* plants photographed under ultraviolet light. Disruption of *MED5a/5b* does not restore normal blue epidermal fluorescence in either *ref8-1* or *ref8-2* mutant plants. **c**, Plants with the same genotypes as **b** at 4.5 weeks after planting, after plants have flowered and begun to set seed.

[1]Department of Biochemistry, Purdue University, West Lafayette, Indiana 47907, USA. [2]Department of Biochemistry, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA. [3]Biosciences Center, National Renewable Energy Laboratory, Golden, Colorado 80401, USA. [4]Department of Agricultural and Biological Engineering and the Laboratory of Renewable Resources Engineering, Purdue University, West Lafayette, Indiana 47907, USA. [5]Department of Biological Systems Engineering, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA. [6]DOE Great Lakes Bioenergy Research Center, and Wisconsin Energy Institute, University of Wisconsin-Madison, Madison, Wisconsin 53726, USA. [7]Weldon School of Biomedical Engineering, Purdue University, West Lafayette, Indiana 47907, USA. †Present addresses: Dow AgroSciences LLC, 9330 Zionsville Road, Indianapolis, Indiana 46268, USA (N.D.B.); Department of Agronomy, University of Wisconsin-Madison, 1575 Linden Drive, Madison, Wisconsin 53706, USA (J.M.).

Plants containing an inactivating insertional mutation in the *REF8* gene halt development after the production of only two or three extremely small leaves[15,16]. Despite retaining all of the phenylpropanoid pathway enzymes required for the synthesis of H lignin subunits, *ref8* mutants synthesize substantially reduced levels of total lignin[10,16], suggesting that lignin biosynthesis may be actively repressed in response to reductions in C3′H activity.

The Mediator complex subunits MED5a and MED5b have recently been shown to be required for homeostatic repression of phenylpropanoid biosynthesis in *Arabidopsis*[17]. Mediator is a large, multisubunit transcriptional co-regulator that is increasingly recognized as an essential element of both basal and regulated eukaryotic transcription[18]. Whereas some subunits of Mediator are required globally for transcription, other subunits are dispensable and appear to be required only for specific transcriptionally regulated processes. In plants, these processes include pathogen response[19,20], freezing tolerance[21], transition to reproductive growth[22] and phytochrome signalling[23].

To test whether MED5a/5b could be involved in the repression of lignin biosynthesis in *ref8-1* mutants, we generated *med5a/5b ref8-1* triple mutants. Although *ref8-1* mutants contain only 40% of the amount of lignin found in wild-type plants, lignin deposition in *med5a/5b ref8-1* triple mutants is restored to wild-type levels (Table 1). In addition, disruption of *MED5a/5b* in the *ref8-1* mutant background led to almost complete rescue of the *ref8-1* growth phenotype, with the morphology of *med5a/5b ref8-1* plants virtually indistinguishable from that of *med5a/5b* plants at multiple stages of growth (Fig. 1b, c). Unlike *ref8-1* mutant plants, which produce stunted inflorescence stems and are uniformly sterile, *med5a/5b ref8-1* plants flower normally and produce robust inflorescence stems and siliques containing fertile seeds (Fig. 1c). Disruption of *MED5a/5b* also ameliorates the developmental arrest phenotype of the null insertional mutant *ref8-2*. Notably, the rescue of *med5a/5b ref8-2* plants is less complete than that of *med5a/5b ref8-1* plants, suggesting that the residual C3′H activity of the latter may be important for their relatively normal growth. On the basis of these results, we conclude that MED5a/5b, and by extension Mediator, has an active role in generating both the lignin deficiency and the dwarf phenotype of *Arabidopsis ref8* mutants.

Observation of *med5a/5b ref8-1* and *med5a/5b ref8-1* mutants under ultraviolet light revealed that they retain the reduced epidermal fluorescence phenotype associated with decreased sinapoylmalate accumulation[10,14] (Fig. 1b), suggesting that the rescue of their stunted growth does not arise from suppression of their underlying metabolic defect. This conclusion is supported by metabolite analysis of *med5a/5b ref8-1* and *med5a/5b ref8-2* mutants, both of which show substantially lower levels of coniferyl and sinapyl alcohol-derived metabolites than wild-type plants, as well as hyperaccumulation of *p*-coumarate and *p*-coumaryl alcohol-derived compounds (Extended Data Fig. 1). The flavonoid hyperaccumulation phenotype associated with loss of C3′H activity[10] is substantially relieved in *med5a/5b ref8-1* and *med5a/5b ref8-2* mutants (Extended Data Fig. 2), consistent with the notion that hyperaccumulation of these compounds is a result of stress-induced synthesis, and not ‘metabolic overflow’, as previously suggested[15,24].

Microscopic and histochemical analysis of inflorescence stem cross-sections revealed multiple cell wall abnormalities in *ref8* mutants that were rescued by disruption of *MED5a/5b*, as well as a number of significant differences between *med5a/5b ref8-1* mutants and wild-type plants. For instance, *med5a/5b ref8-1* cross-sections show significantly more intense staining compared with *ref8-1* cross-sections when treated with either Mäule reagent (Fig. 2a) or phloroglucinol (Fig. 2b), two commonly used stains for lignified tissues. The red staining of S lignin subunits by Mäule reagent, however, is substantially less intense in *med5a/5b ref8-1* mutants than in wild-type or *med5a/5b* mutant plants, consistent with a loss of S subunits due to decreased C3′H activity. Treatment with phloroglucinol, which is thought to react primarily with lignin-associated hydroxycinnamaldehyde moieties[25], also results in more intense staining in *med5a/5b ref8-1* plants than in *ref8-1* plants. The reproducible difference in the colour of phloroglucinol-stained *med5a/5b ref8-1* tissues (bright, reddish-purple) compared with wild-type and *med5a/5b* tissues (darker purple) was also observed in *med5a/5b ref8-2* tissues (Extended Data Fig. 3) and is possibly due to differences in the reaction of phloroglucinol with *p*-coumaraldehyde (4-hydroxycinnamaldehyde) versus coniferaldehyde (4-hydroxy-3-methoxycinnamaldehyde). Vascular bundles of *med5a/5b ref8-1* plants contain relatively normal, open xylem elements, with only a few exhibiting partial collapse, in contrast with those of *ref8-1* plants, which uniformly exhibit severely collapsed and irregular xylem (Fig. 2a, b). At higher levels of magnification, using either confocal (Fig. 2c) or transmission electron microscopy (Fig. 2d), secondary cell walls of *ref8-1* mutants show considerable disorganization and are significantly thicker (Extended Data Fig. 4) than wild-type cell walls, possibly due to altered hydrophobicity and architecture of the wall in the absence of normal levels of lignin. By contrast, cell walls of *med5a/5b ref8-1* mutants show relatively normal architecture, but are slightly thinner than those of wild-type plants (Fig. 2d and Extended Data Fig. 4).

To obtain quantitative information on the lignin composition and substructure of *med5a/5b ref8-1* cell walls, we used solution-state two-dimensional nuclear magnetic resonance spectroscopy (2D-NMR) and derivatization followed by reductive cleavage (DFRC) lignin analyses.

Table 1 | Thioglycolic acid quantification of total lignin

Genotype	$A_{280\,nm}$ mg fresh weight^{-1} ± s.d.*
Wild type	0.116 ± 0.010
med5a/5b	0.122 ± 0.007
ref8-1	0.046 ± 0.003
med5a/5b ref8-1	0.102 ± 0.011

$A_{280\,nm}$, absorbance at 280 nm.
* *N* = 3.

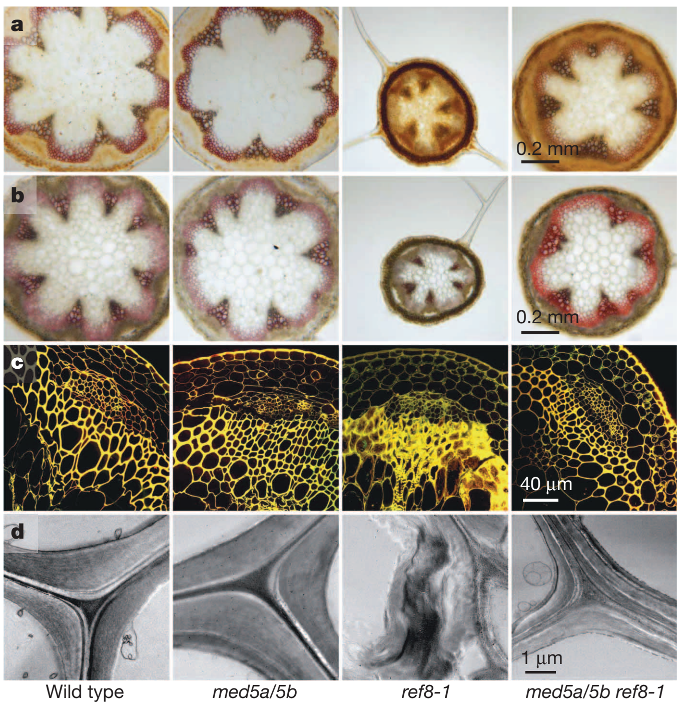


Figure 2 | Secondary cell wall composition and architecture of *med5a/5b ref8-1* mutant plants is distinct from both wild-type and *ref8-1* mutant plants. **a**, Mäule staining, which stains S subunits red and G subunits orange-brown. **b**, Phloroglucinol staining, which is specific for lignin-associated hydroxycinnamaldehydes. **c**, Confocal microscopy of tissue stained with the fluorescent dye acriflavin. **d**, Transmission electron microscopy of cell wall corners between adjacent fibre cells. All images are derived from thin sections of inflorescence stems of wild-type, *med5a/5b*, *ref8-1* and *med5a/5b ref8-1* plants, as indicated at the bottom of each column.

Examination of the aromatic regions of short-range ^{1}H–^{13}C correlation spectra revealed that lignin isolated from *med5a/5b ref8-1* mutants consists almost entirely (~95%) of H-lignin subunits, and yields NMR signal patterns nearly identical to those of a synthetic lignin prepared solely from *p*-coumaryl alcohol (Fig. 3a). By contrast, both wild type and *med5a/5b* mutants contain typical G-rich, coniferyl- and sinapyl-alcohol-derived G/S lignin, with H monomers accounting for less than 2% of total lignin. These NMR data are consistent with lignin composition of the same plants as determined by the DFRC method (Fig. 3b), which quantifies lignin monomeric units released by cleavage of aryl-ether-type lignin intersubunit linkages[26]. The high level of H lignin in *med5a/5b ref8-1* mutants is well beyond that previously reported in C3′H (and related hydroxycinnamoyl-CoA shikimate hydroxycinnamoyl-transferase (HCT)) knockdown plants, and importantly is not accompanied by the aberrant cell wall architecture and morphological abnormalities that are typically associated with loss of C3′H (Fig. 2c, d). Also revealed by 2D-NMR were major changes in lignin substructure arising from differences in the relative abundance of the various intersubunit linkages (Extended Data Fig. 5). Most notably, *med5a/5b ref8-1* plants showed a substantial decrease in aryl-ether-type (β–O–4) linkages that arise from the addition of a lignin monomer to the phenolic end of a growing polymer and are the most abundant in wild-type lignin[7]. This decrease is offset by a twofold increase in phenylcoumaran-type (β–5) and a threefold increase in resinol-type (β–β) linkages. These changes in turn provide an explanation for the significant reduction in total lignin monomers determined by the DFRC method (Fig. 3b), as this method does not release lignin monomers linked to adjacent monomers via either β–5 or β–β linkages[26]. We also observed a substantial (~16-fold) increase in the proportion of normally minor arylglycerol end units in the high H lignin of *med5a/5b ref8-1* mutants. Similar changes in lignin substructures were previously observed during our NMR analyses of other high-H-lignin plants[16,27] and are probably attributable to the substantial differences in the polymerization characteristics of *p*-coumaryl alcohol versus the typical coniferyl and sinapyl alcohols[27]. The observation of elevated proportions of arylglycerol end units also suggests that the lignin of *med5a/5b ref8-1* mutants consists of a greater number of shorter polymers (that is, exhibits a lower degree of polymerization) than the lignin of wild-type and *med5a/5b* plants. Consistent with this interpretation, gel-permeation chromatography analyses revealed a significant under-representation of high-molecular-weight species in lignin of *med5a/5b ref8-1* plants compared with wild-type and *med5a/5b* lignin (Extended Data Fig. 6).

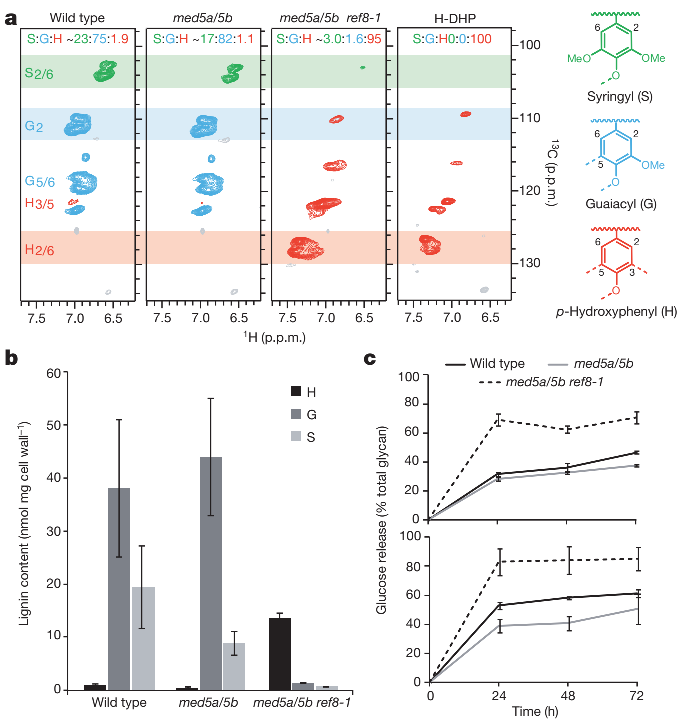


Figure 3 | Lignocellulosic material from *med5a/5b ref8-1* plants contains lignin composed almost entirely of H subunits and shows substantially increased saccharification potential. **a**, 2D-NMR analysis of lignin composition of wild-type, *med5a/5b* and *med5a/5b ref8-1* plants, and a synthetic H-only lignin (H-dehydrogenation polymer (H-DHP)). Resonance signals arising from H, G and S subunits are colour coded to match the structures shown on the right. **b**, Gas chromatographic quantification of H, G and S lignin monomers after reductive cleavage of aryl-ether linkages by the DFRC method. **c**, Glucose yield after incubation of lignocellulosic material with cellulase and β-glucosidase either without pre-treatment (top) or after liquid hot water pre-treatment (bottom). Error bars in **b** and **c** represent the standard deviation (s.d.) of three biological replicates.

To test whether the altered lignin composition and structure of *med5a/5b ref8-1* plants leads to facilitated breakdown of cell wall polysaccharides, lignocellulosic material from mutant and wild-type plants was subjected to a mixture of commercial cellulase and β-glucosidase enzymes and the resulting release of glucose monitored over time (Fig. 3c). These analyses revealed a substantial increase in glucose yield from *med5a/5b ref8-1* biomass compared with samples from wild-type or *med5a/5b* plants. Non-pre-treated *med5a/5b ref8-1* samples, for instance, showed more than twice the glucose yield of non-pre-treated wild-type and *med5a/5b* samples after 24 h of treatment with cellulase and β-glucosidase (~70% versus ~30% of total glycan). Pre-treatment of *med5a/5b ref8-1* material with liquid hot water increased the glucose yield to ~80% of total glycan after 24 h, compared with 50% and 40% of total glycan for wild-type and *med5a/5b* material, respectively. In the case of both pre-treated and non-pre-treated *med5a/5b ref8-1* material, increasing the time of cellulase/β-glucosidase incubation did not significantly increase the total glucose yield. These data suggest that genetically modified, high-H-lignin feedstocks could potentially be used to reduce the cost of biomass pre-treatment and enzymatic saccharification, thus substantially improving the economic balance sheet of lignocellulosic biofuel production.

To determine the possible effects on gene expression of disrupting *MED5a/5b* in the *ref8-1* mutant background, we used high-throughput messenger RNA sequencing to identify differentially expressed genes among wild-type, *med5a/5b*, *ref8-1* and *med5a/5b ref8-1* plants (Fig. 4 and Supplementary Data 1). Consistent with our previous report[17], *med5a/5b* plants exhibited elevated transcript levels of phenylpropanoid biosynthetic genes, including those encoding PAL, C4H, 4CL, C3′H, CCR, CAD and the recently identified[28] caffeoylshikimate esterase, CSE (Extended Data Fig. 7). Moreover, ‘phenylpropanoid biosynthesis’ was the term most significantly overrepresented among ontologies of the 248 genes overexpressed in *med5a/5b* plants (Supplementary Table 1). In *ref8-1* plants, a total of 8,772 genes, representing a substantial fraction of the genome, were expressed at levels significantly different from wild type. Drought- and dehydration-responsive genes showed the most significant association with increased expression in *ref8-1* plants (Supplementary Table 2), consistent with the suggestion that lignin deficiency may lead to impaired water transport. The increased expression of flavonoid biosynthesis genes in *ref8-1* plants (Extended Data Fig. 8) further supports the conclusion that flavonoid hyperaccumulation in this mutant is due to stress-induced synthesis rather than metabolic overflow. Also among the genes overexpressed in *ref8-1* mutants were the same phenylpropanoid genes that showed increased expression in *med5a/5b* mutants (Extended Data Fig. 7), indicating that the reduced total lignin in *ref8-1* mutants is not a result of Mediator-dependent transcriptional repression of these genes. Consistent with the reduced size of *ref8-1* mutants, gene ontology term analysis revealed that a high proportion of the genes with decreased expression in these plants were related to basic processes of growth, cell division and primary metabolism (Supplementary Table 3). Remarkably, of the ~9,000 genes exhibiting increased or decreased transcript levels in *ref8-1* plants, nearly 90% were expressed at wild-type levels in *med5a/5b ref8-1* plants (Fig. 4a). Of the genes exhibiting differential expression in both *ref8-1* and *med5a/5b ref8-1* plants, a substantial

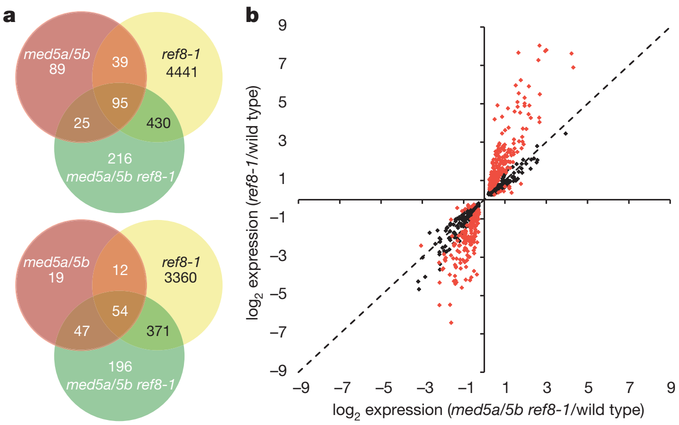


Figure 4 | Disruption of *MED5a/5b* rescues the widespread transcriptional reprogramming of the *ref8-1* mutant. **a**, Venn diagrams showing the number of genes with significantly increased (top) or decreased (bottom) expression in *med5a/5b*, *ref8-1* and *med5a/5b ref8-1* mutants compared with wild-type plants, as well as the intersection of these sets. **b**, Relationship between expression in *med5a/5b ref8-1* mutants (*x*-axis) and *ref8-1* mutants (*y*-axis) of all genes either upregulated (upper right quadrant) or downregulated (lower left quadrant) in both *ref8-1* and *med5a/5b ref8-1* plants compared with wild type. Each gene is represented by one point. Genes exhibiting a significant difference in expression level between *ref8-1* and *med5a/5b ref8-1* are shown in red. For all RNA-seq data, significance corresponds to a multiple-test-adjusted P value < 0.05, as determined using the DESeq algorithm. Genes for which there is no significant difference are shown in black.

number are misregulated to a lesser degree in *med5a/5b ref8-1* plants, with transcript levels much closer to those in wild-type plants (Fig. 4b). With the exception of MYB75 (also known as PAP1) and MYB90 (also known as PAP2), the expression of phenylpropanoid-controlling transcription factors is largely unaffected in the mutants. The moderate overexpression of *MYB75* and *MYB90* in *ref8-1* and *med5a/5b ref8-1* mutants is consistent with their elevated levels of anthocyanins, but is unlikely to contribute to their dwarf phenotype, as the *pap1-D* mutant is not dwarfed, despite much greater MYB75 overexpression[29]. Taken together, these results demonstrate that genetic disruption of C3′H activity results in widespread qualitative and quantitative changes in gene expression in *ref8-1* mutant plants, and that MED5a/5b are required, directly or indirectly, for the great majority of these changes.

The molecular events underlying the dwarfing of lignin-deficient mutants are complex and poorly understood[6]. The experiments described here provide new insights into this phenomenon by demonstrating that MED5a/5b are required for both the stunted growth and the lignin deficiency of *Arabidopsis ref8* mutants. Taken together with the established roles of the Mediator complex in transcriptional regulation, and of MED5a/5b in the repression of phenylpropanoid metabolism, these observations lead us to propose the following model (Extended Data Fig. 9). MED5a/5b are key components of an active, transcriptional process by which phenylpropanoid homeostasis is maintained in wild-type plants. In *ref8* mutants, the metabolic block at C3′H leads to changes in one or more wall-bound or soluble phenylpropanoid metabolites, eliciting an inappropriate or exaggerated response by this normally homeostatic pathway and initiating a transcriptional cascade that ultimately results in repression of lignification and impaired growth. The failure to induce or repress one or more direct targets of MED5a/5b in *med5a/5b ref8* plants interrupts this response, thereby restoring their ability to grow and to synthesize wild-type levels of a novel, nearly pure H lignin.

The surprising observation that G and S lignin are largely dispensable for normal growth and development in the permissive *med5a/5b* background raises important new questions for future investigation. Given that *ref8* mutants do not exhibit transcriptional repression of phenylpropanoid biosynthetic genes, and that H subunits can be polymerized into a functional lignin polymer, it is unclear why *ref8* mutants should fail to lignify normally. One possibility is that, in addition to their role in transcriptional regulation of phenylpropanoid biosynthesis, MED5a/5b may also control the expression of one or more target genes responsible for post-transcriptional repression of monolignol synthesis, transport or polymerization. If true, this would suggest that the failure to produce sufficient total lignin, and not the loss of G and S subunits, could be the underlying cause of dwarfing in *ref8* mutants. Consistent with this possibility, we have observed that disruption of *MED5a/5b* does not rescue the stunted growth of the C4H-deficient *ref3-2* mutant (data not shown), indicating that alterations in the Mediator complex do not relieve the requirement for an unobstructed pathway to at least one of the canonical monolignols. Alternatively, it is possible that Mediator represses the growth of *ref8* mutants independent of its role in repressing lignification, for instance by modulating the activity of a wall integrity-sensing pathway similar to those mediated by THE1 or WAK1. Indeed, we cannot formally exclude the possibility that repression of lignification in *ref8* mutants is the result, rather than the cause, of their impaired growth. It will be important to identify the direct targets of MED5a/5b among the thousands of misregulated genes in *ref8* mutants to distinguish between these alternative models. It is also notable that although it has been suggested that dwarfing in lignin-deficient mutants is due to hyperaccumulation of salicylic acid[30], this is apparently not the case for *Arabidopsis ref8-1* mutants; rescued *med5a/5b ref8-1* mutants contain higher levels of salicylic acid than *ref8-1* plants (Extended Data Fig. 10a), and disruption of the salicylic acid biosynthetic enzyme isochorismate synthase does not alleviate the dwarfism of *ref8-1* mutants (Extended Data Fig. 10b).

Finally, our saccharification analysis of *med5a/5b ref8-1* plants demonstrates that in the proper genetic background it is possible to create high-H-lignin plants with greatly reduced biomass recalcitrance, while at the same time avoiding the severe yield penalty normally associated with loss of G and S lignin. Together with previously reported evidence that the role of Mediator in phenylpropanoid repression may be conserved[17], this observation suggests that similar genetic modifications could lead to improved biomass crops for forage and bioenergy. Further elucidation of the gene products upstream and downstream of Mediator in the transcriptional response to phenylpropanoid pathway defects promises to yield further insight into this phenomenon and may reveal additional targets for bioengineering.

METHODS SUMMARY

Plants simultaneously containing mutations in *MED5a*, *MED5b* and *REF8* were generated by crossing previously described transfer DNA (T-DNA)- and ethyl methanesulphonate (EMS)-derived mutants. PCR-based methods were used for genotype determination and for the identification of mutants of interest. Cell walls of mutant and control plants were visualized using a variety of staining methods in conjunction with bright-field, confocal and transmission electron microscopy. Lignin composition and substructures were determined using 2D-NMR as well as DFRC, a gas-chromatography-based method that detects lignin monomers liberated by cleavage of all lignin β–O–4 bonds. Changes in soluble phenylpropanoid metabolism were monitored using high-performance liquid chromatography. Recalcitrance of lignocellulose was determined by treating with commercial cellulase and β-glucosidase, and monitoring the release of glucose over time. Gene expression differences between genotypes were determined by high-throughput sequencing of mRNA extracted from whole rosettes before the transition to flowering (that is, RNA-seq).

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 22 March 2013; accepted 27 January 2014.
Published online 16 March 2014.

1. Bonawitz, N. D. & Chapple, C. The genetics of lignin biosynthesis: connecting genotype to phenotype. *Annu. Rev. Genet.* **44,** 337–363 (2010).
2. Vanholme, R. *et al.* Lignin biosynthesis and structure. *Plant Physiol.* **153,** 895–905 (2010).
3. Bonawitz, N. D. & Chapple, C. Can genetic engineering of lignin biosynthesis be accomplished without an unacceptable yield penalty? *Curr. Opin. Biotechnol.* **24,** 336–343 (2013).

4. Rogers, L. A. & Campbell, M. M. The genetic control of lignin deposition during plant growth and development. *New Phytol.* **164,** 17–30 (2004).
5. Chen, F. & Dixon, R. A. Lignin modification improves fermentable sugar yields for biofuel production. *Nature Biotechnol.* **25,** 759–761 (2007).
6. Hématy, K. *et al.* A receptor-like kinase mediates the response of *Arabidopsis* cells to the inhibition of cellulose synthesis. *Curr. Biol.* **17,** 922–931 (2007).
7. Ralph, J. *et al.* Lignins: natural polymers from oxidative coupling of 4-hydroxyphenylpropanoids. *Phytochem. Rev.* **3,** 29–60 (2004).
8. Chapple, C. C., Vogt, T., Ellis, B. E. & Somerville, C. R. An *Arabidopsis* mutant defective in the general phenylpropanoid pathway. *Plant Cell* **4,** 1413–1424 (1992).
9. Franke, R. *et al.* Changes in secondary metabolism and deposition of an unusual lignin in the *ref8* mutant of *Arabidopsis. Plant J.* **30,** 47–59 (2002).
10. Franke, R. *et al.* The *Arabidopsis REF8* gene encodes the 3-hydroxylase of phenylpropanoid metabolism. *Plant J.* **30,** 33–45 (2002).
11. Hoffmann, L. *et al.* Silencing of hydroxycinnamoyl-coenzyme A shikimate/quinate hydroxycinnamoyltransferase affects phenylpropanoid biosynthesis. *Plant Cell* **16,** 1446–1465 (2004).
12. Schilmiller, A. L. *et al.* Mutations in the cinnamate 4-hydroxylase gene impact metabolism, growth and development in *Arabidopsis. Plant J.* **60,** 771–782 (2009).
13. Reddy, M. S. *et al.* Targeted down-regulation of cytochrome P450 enzymes for forage quality improvement in alfalfa (*Medicago sativa* L.). *Proc. Natl Acad. Sci. USA* **102,** 16573–16578 (2005).
14. Ruegger, M. & Chapple, C. Mutations that reduce sinapoylmalate accumulation in *Arabidopsis thaliana* define loci with diverse roles in phenylpropanoid metabolism. *Genetics* **159,** 1741–1749 (2001).
15. Abdulrazzak, N. *et al.* A coumaroyl-ester-3-hydroxylase insertion mutant reveals the existence of nonredundant *meta*-hydroxylation pathways and essential roles for phenolic precursors in cell expansion and plant growth. *Plant Physiol.* **140,** 30–48 (2006).
16. Weng, J. K. *et al.* Convergent evolution of syringyl lignin biosynthesis via distinct biosynthetic pathways in the lycophyte *Selaginella* and flowering plants. *Plant Cell* **22,** 1033–1045 (2010).
17. Bonawitz, N. D. *et al.* REF4 and RFR1, subunits of the transcriptional coregulatory complex Mediator, are required for phenylpropanoid homeostasis in *Arabidopsis. J. Biol. Chem.* **287,** 5434–5445 (2012).
18. Kornberg, R. D. The molecular basis of eukaryotic transcription. *Proc. Natl Acad. Sci. USA* **104,** 12955–12961 (2007).
19. Kidd, B. N. *et al.* The Mediator complex subunit PFT1 is a key regulator of jasmonate-dependent defense in *Arabidopsis. Plant Cell* **21,** 2237–2252 (2009).
20. Wathugala, D. L. *et al.* The Mediator subunit SFR6/MED16 controls defence gene expression mediated by salicylic acid and jasmonate responsive pathways. *New Phytol.* **195,** 217–230 (2012).
21. Knight, H. *et al.* Identification of SFR6, a key component in cold acclimation acting post-translationally on CBF function. *Plant J.* **58,** 97–108 (2009).
22. Imura, Y. *et al.* CRYPTIC PRECOCIOUS/MED12 is a novel flowering regulator with multiple target steps in *Arabidopsis. Plant Cell Physiol.* **53,** 287–303 (2012).
23. Cerdán, P. D. & Chory, J. Regulation of flowering time by light quality. *Nature* **423,** 881–885 (2003).
24. Besseau, S. *et al.* Flavonoid accumulation in *Arabidopsis* repressed in lignin synthesis affects auxin transport and plant growth. *Plant Cell* **19,** 148–162 (2007).
25. Clifford, M. N. Specificity of acidic phloroglucinol reagents. *J. Chromatogr. A* **94,** 321–324 (1974).
26. Lu, F. & Ralph, J. The DFRC method for lignin analysis. 2. Monomers from isolated lignins. *J. Agric. Food Chem.* **46,** 547–552 (1998).
27. Ralph, J. *et al.* Effects of coumarate 3-hydroxylase down-regulation on lignin structure. *J. Biol. Chem.* **281,** 8843–8853 (2006).
28. Vanholme, R. *et al.* Caffeoyl shikimate esterase is an enzyme in the lignin biosynthetic pathway in *Arabidopsis. Science* **341,** 1103–1106 (2013).
29. Borevitz, J. O., Xia, Y., Blount, J., Dixon, R. A. & Lamb, C. Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. *Plant Cell* **12,** 2383–2394 (2000).
30. Gallego-Giraldo, L., Escamilla-Trevino, L., Jackson, L. A. & Dixon, R. A. Salicylic acid mediates the reduced growth of lignin down-regulated plants. *Proc. Natl Acad. Sci. USA* **108,** 20814–20819 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was primarily funded by the Division of Chemical Sciences, Geosciences, and Biosciences, Office of Basic Energy Sciences of the US Department of Energy (DOE) through grant DE-FG02-07ER15905 to C.C. N.D.B. was supported in part by a fellowship from the Life Sciences Research Foundation. Y.T. and J.R. were funded by the US DOE Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-FC02-07ER64944). J.I.K., P.N.C. and B.S.D. were supported as part of the Center for Direct Catalytic Conversion of Biomass to Biofuels (C3Bio), an Energy Frontier Research Center funded by the US DOE, Office of Science, Office of Basic Energy Sciences, award number DE-SC0000997. E.X. and M.L. were supported by the US DOE through grant DE-FG02-06ER64301 to M.L. and C.C. and by the Purdue University Office of Agricultural Research Programs. The authors acknowledge the support of the Bioinformatics Core at Purdue University.

Author Contributions N.D.B. and C.C. were responsible for the conception, planning and organization of experiments. N.D.B. generated all *Arabidopsis* lines and all plant material used in the work, carried out metabolite analysis, interpretation of RNA-seq results and TGA lignin quantification, and was responsible for photography of whole plants. N.D.B. and J.I.K. harvested samples and prepared RNA for global transcript analysis. J.I.K. performed histochemical analyses and quantification of salicylic acid. N.A.A. assisted with cell-wall-bound phenylpropanoid quantification. J.M. carried out DFRC lignin analysis. P.N.C. and B.S.D. performed confocal microscopy, transmission electron microscopy, and cell wall thickness measurements. Y.T. and J.R. were responsible for NMR and gel-permeation chromatography analysis of lignin. E.X. and M.L. performed saccharification assays. The manuscript was primarily written by N.D.B. with critical input from other co-authors. Figures were prepared by N.D.B. with support from Y.T. and P.N.C.

Author Information RNA-seq data have been deposited in the NCBI Sequence Read Archive under accession number SRP035272. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.C. (chapple@purdue.edu).

METHODS

Plant material and growth conditions. Plants were grown in Redi-earth Plug and Seedling Mix (Sun Gro Horticulture) supplemented with Scotts Osmocote Plus controlled release fertilizer (Hummert International) at a temperature of 22 °C and a light intensity of $100\,\mu E\,m^{-2}\,s^{-1}$, with a 16 h light/8 h dark photoperiod. All photographic images presented are unaltered with respect to brightness, colour and contrast. In some cases, portions of single images have been cropped or rearranged for neatness and clarity of presentation.

The *Arabidopsis ref8-1* and *ref3-2* mutants were isolated from a forward genetic screen for plants deficient in the synthesis of sinapoylmalate and have been described previously[10,11,13,15]. The *sid2-4* mutant (The *Arabidopsis* Information Resource (TAIR) accession number SALK_133146) contains a transfer DNA (T-DNA) insertion in the second intron of the gene encoding the salicylic acid biosynthetic enzyme isochorismate synthase and was obtained from the *Arabidopsis* Biological Research Center (ABRC; The Ohio State University). Primers CC4042 (left primer, 5′-TCT GGGCTCAAACACTAAAACAC-3′), CC4043 (right primer, 5′-GAATCAGAG GTGACGTTGAAGAC-3′) and CC2449 (border primer, 5′-ATTTTGCCGATT TCGGAAC-3′) were used to genotype plants for the *sid2-4* insertion. To generate the *ref8-1 sid2-4* mutant, we began by crossing a *sid2-4* homozygous plant to a confirmed *REF8-1/ref8-1* heterozygote. Plants heterozygous for the *ref8-1* mutation among the F1 progeny of this cross were identified as previously described[17], and allowed to self-fertilize. In the resulting population, we identified plants with the genotype *REF8/ref8-1 sid2-4/sid2-4*, which give rise to progeny with the genotypes *REF8/REF8 sid2-4/sid2-4*, *REF8/ref8-1 sid2-4/sid2-4*, and *ref8-1/ref8-1 sid2-4/sid2-4* at a ratio of 1:2:1. The *ref8-2* (TAIR accession SALK_036132) mutant, as well as the *med5a* (TAIR accession SALK_037472) and *med5b* (TAIR accession SALK_011621) mutants used to generate the *med5a/5b* double mutant were originally obtained from the ABRC and have also been previously described[17,18,31]. To generate the *med5a/5b ref8-1* triple mutant, we began by crossing a *med5a/5b* mutant plant to a confirmed *REF8/ref8-1* heterozygote. Plants heterozygous for the *ref8-1* mutation among the F1 progeny of this cross were identified and allowed to self-fertilize as described earlier. We then used high-performance liquid chromatography (HPLC) to screen F2 progeny of this self-fertilization for plants exhibiting the characteristic metabolite profile of *med5a/5b* homozygous double mutants. These plants were then genotyped using PCR as previously described[18] to identify plants with the genotype *med5a/med5a med5b/med5b REF8/ref8-1*. These plants were again allowed to self-fertilize, giving rise to *med5a/med5a med5b/med5b REF8/REF8*, *med5a/med5a med5b/med5b REF8/ref8-1*, and *med5a/med5a med5b/med5b ref8-1/ref8-1* plants in the F3 generation at a ratio of 1:2:1. An identical approach was used to generate *med5a/5b ref8-2* and *med5a/5b ref3-2* plants. To genotype for the presence of the *ref3-2* allele, the primer pair CC2396 (5′-TTCCG TATCATGTTCGATAG-3′) and CC2397 (5′- AATGTCAATTTCCCAAAATC-3′) was used in combination with HinFI digestion, exploiting the cleaved amplified polymorphic sequence marker resulting from the point mutation in *ref3-2*. Genotyping for the *ref8-2* allele has been described previously[17]. The wild type and all mutants are in the Columbia-0 (Col-0) background.

Histochemical staining. For histochemical staining, basal sections of primary inflorescence stems were embedded in 5% agar and cut into 100 μm sections using a vibratome. For phloroglucinol staining, sections were then incubated in 16% ethanol (v/v), 10% HCl (v/v), 0.2% phloroglucinol (w/v) for 10 min, then washed, mounted in water, and examined using an Olympus Vanox-S light microscope. For Mäule staining, sections were fixed in 4% glutaraldehyde, then washed with water, incubated in 0.5% $KMnO_4$ (w/v) for 5 min and washed twice with water. Finally, sections were mounted in concentrated NH_4OH and examined under the light microscope.

Confocal laser scanning microscopy. Semi-thin sectioned samples were positioned on glass microscope slides and stained with 0.1% acriflavin. Images were captured using a Nikon C1 Plus microscope, equipped with the Nikon C1 confocal system with four lasers (403 nm, 561 nm, 643 nm and Argon tuneable 458/477/488/515 nm), and operated via Nikon's EZ-C1 software.

Cell wall thickness measurements. Cell wall thickness was measured directly from the confocal laser scanning microscopy (CLSM) images using tools within the Fiji (fiji.sc) image processing package. The high contrast of the acriflavin-stained 300-nm-thick sections imaged by CLSM allowed images of a group of similar cells such as vascular bundle fibre cells to be isolated as a region of interest (ROI), thresholded, and converted to binary easily and accurately (Extended Data Fig. 4). From the binary image, two operations were performed to derive a description of cell wall thickness. First, skeletonization provides a set of coordinates of the midpoint of the cell walls. Second, a distance map provides the Euclidian distance of each point within the cell wall to the nearest non-cell-wall space. By combining these two, the distance measurement is reported for each point along the cell wall.

Transmission electron microscopy. Stem sections were high-pressure frozen with a Leica EMPact2 high-pressure freezer in 0.2 mm deep planchets (Leica Microsystems) using 0.15 M sucrose as a cryoprotectant. Next, freeze substitution was carried out in a Leica AFS2 automated freeze substitution unit in 1% OsO_4 (EMS). The samples were dehydrated by solution exchange with increasing concentrations of acetone. After dehydration, the samples were infiltrated with Eponate 812 (EMS) by incubating at room temperature for several hours to overnight in increasing concentrations of resin diluted in acetone. The samples were transferred to capsules and the resin polymerized in a 60 °C oven overnight. Resin-embedded samples were sectioned to ~50 nm with a Diatome diamond knife on a Leica EM UTC ultramicrotome (Leica Microsystems). Sections were collected on 0.5% Formvar-coated slot grids (SPI Supplies) and were post-stained for 1 min with 1% aqueous $KMnO_4$. Images were taken with a 4 mega-pixel Gatan UltraScan 1000 camera on a FEI Tecnai G2 20 Twin 200 kV LaB6 transmission electron microscope.

Plant cell wall preparation. Extractive-free plant cell walls for DFRC, NMR, and saccharification analyses were prepared essentially as previously described[32]. Briefly, mature, dried *Arabidopsis* inflorescence stems were stripped of all cauline leaves and siliques, cut into small pieces, and finely ground in liquid nitrogen. Ground tissue was then serially extracted in ten volumes of 0.1 M sodium phosphate buffer (pH 7.2) at 50 °C for 1 h, six extractions in ten volumes of 70% ethanol (v/v) at 70 °C for 15 min, and a final extraction in 100% acetone for 10 min. Cell wall residue was then collected by centrifugation and dried overnight in a vacuum oven at 50 °C.

Thioglycolic acid quantification of lignin. For thioglycolic acid quantification of lignin, approximately 100 mg of fresh inflorescence stem material was flash-frozen in liquid nitrogen and ground in a microcentrifuge tube with a small pestle. Ground tissue was then extracted in ten volumes of 100% methanol at 80 °C for 2 h, collected by centrifugation, washed with ten volumes of distilled water, and again collected by centrifugation. Samples were then resuspended in 750 μl of distilled water, 250 μl concentrated HCl and 100 μl thioglycolic acid and incubated for 3 h at 80 °C. Samples were then collected by centrifugation, washed with 1 ml of distilled water, resuspended in 1 ml of 1 M NaOH, and incubated with gentle shaking for 12 h at room temperature. After spinning for 10 min at maximum speed in a microcentrifuge, the supernatant was transferred to a new tube, to which was added 200 μl of concentrated HCl. Samples were then vortexed and incubated at 4 °C for 4 h. The precipitate of this reaction was then collected by spinning for 10 min at maximum speed in a microcentrifuge, dissolved in 1 ml of 1 M NaOH, and the absorbance at 280 nm determined using a spectrophotometer.

DFRC. DFRC lignin analysis was performed essentially as previously reported[26]. Briefly, cell wall samples were dissolved in acetyl bromide/acetic acid solution, containing 4,4′-ethylidenebisphenol as an internal standard. The reaction products were dried down using nitrogen gas, dissolved in dioxane/acetic acid/water (5/4/1, v/v/v), reacted with Zn dust, purified with C-18 SPE columns (Supelco), and acetylated with pyridine/acetic anhydride (2/3, v/v). The lignin derivatives were analysed by gas chromatography/flame ionization detection using response factors relative to the internal standard of 1.26 for *p*-coumaryl alcohol peracetate, 1.30 for coniferyl alcohol peracetate and 1.44 for sinapyl alcohol peracetate. The same samples were run through gas chromatography–mass spectrometry in parallel to confirm the identity of the derived hydroxycinnamyl alcohol peracetates.

Preparation of cellulolytic enzyme lignins from *Arabidopsis* stem cell walls. Preparation of cellulolytic enzyme lignin (CEL) samples from *Arabidopsis* stem cell walls for NMR and gel-permeation chromatography (GPC) was as described previously[33,34]. In brief, preground cell wall samples were extracted with 80% aqueous ethanol (sonication 3 × 20 min). Isolated cell walls (~300 mg) were ball milled (7 × 5 min milling and 5 min cooling cycles) using a Retsch PM100 ball mill vibrating at 600 r.p.m. with ZrO_2 vessels containing ZrO_2 ball bearings. The ball-milled walls (~250 mg) were transferred to centrifuge tubes and digested at 30 °C with crude cellulases (Calbiochem Cellulysin; lot no. D00074989; $30\,mg\,g^{-1}$ of sample in pH 5.0 acetate buffer; three times over 2 days, fresh buffer and enzyme added each time), leaving all of the lignin and residual polysaccharides (21%, 24% and 20% isolation yields for the original cell walls of wild-type, *med5a/5b* and *med5a/5b ref8-1* plants). For NMR and GPC analysis, the cellulase-digested cell walls (~35 mg) were subjected to solubilization and acetylation in DMSO/*N*-methylimidazole/acetic anhydride[35] to afford acetylated CELs (137%, 142% and 141% weight yields for CELs from wild-type, *med5a/5b* and *med5a/5b ref8-1* plants).

Preparation of *in vitro* synthetic lignin polymer from *p*-coumaryl alcohol. The *p*-coumaryl alcohol precursor was synthesized as previously described[36]. H-DHP was generated via horseradish peroxidase (HRP)-catalysed polymerization using a typical end-wise polymerization method[37]: 240 ml of acetone/sodium phosphate buffer (0.1 M, pH 6.5) (1:9, vol/vol) containing *p*-coumaryl alcohol (1 mmol), and a separate solution of hydrogen peroxide (1.2 mmol) in 240 ml of water were added by peristaltic pump over a 20 h period at 25 °C to 60 ml of buffer containing HRP (Sigma-Aldrich; type VI, 250–330 U, 5 mg). The reaction mixture was further stirred for 4 h and then acidified to pH ~3 with 1 M aqueous HCl. The precipitate was collected by centrifugation (10,000*g*, 15 min), washed with ultrapure water (100 ml

$\times$ 3), and lyophilized (weight yield 91%). Acetylation of DHPs (~40 mg) used 1:1 acetic anhydride and pyridine (weight yield, 112%).

NMR spectroscopy. NMR spectra were acquired on a Bruker Biospin AVANCE 500 MHz spectrometer fitted with a cryogenically cooled 5 mm TCI gradient probe with inverse geometry (^{1}H coils closest to the sample). Acetylated samples of *Arabidopsis* CELs (~50 mg) or H-DHP (~40 mg) were dissolved in 0.5 ml of chloroform-*d*; the central chloroform solvent peak was used as an internal reference (δ_C, 77.0; δ_H, 7.26 p.p.m.). HSQC experiments using Bruker's adiabatic pulse version of the experiment (hsqcetgpsisp.2) were carried out using the following parameters[33]: acquired from 10 to 0 p.p.m. in F2 (^{1}H) with 1,998 data points (acquisition time 200 ms), 200 to 0 p.p.m. in F1 (^{13}C) with 400 increments (F1 acquisition time 8 ms) of 96 scans with a 1.0 s interscan delay; the d_{24} delay was set to 0.89 ms (1/8J, J: 140 Hz). Processing used typical matched Gaussian apodization in F2 and squared cosine-bell apodization and one level of linear prediction (32 coefficients) in F1. Volume integration of contours in HSQC plots used Bruker's TopSpin 3.1 software and no correction factors were used; that is, the data represent volume integrals only. For quantification of H/G/S distributions, only the C_2–H_2 correlations from G units and the C_2–H_2/C_6–H_6 correlations from H and S units were used, and the G integrals were logically doubled. For rough estimation of the various interunit linkage types, the following well-resolved contours were integrated: $\mathbf{A}_\alpha$, $\mathbf{B}_\alpha$, $\mathbf{C}_\alpha$, $\mathbf{D}_\alpha$, $\mathbf{D'}_\alpha$, $\mathbf{E}_\alpha$, $\mathbf{X1}_\gamma$ and $\mathbf{X2}_\beta$; **X1** and **X2** are not included in the total, which reflects just the interunit linkages; their percentages in Extended Data Fig. 5 are expressed as a percentage of total interunits (**A**–**E**).

HPLC analysis of soluble metabolites. To extract and measure hydroxycinnamate esters and flavonols, 3-week-old rosette leaves were harvested and extracted in 50% methanol (v/v) for 1 h at 65 °C at a tissue concentration of 100 mg ml^{-1}. Insoluble material was sedimented by spinning for 5 min at 14,000 r.p.m. in a microcentrifuge. Ten microlitres of the supernatant was then loaded onto a Shim-pack XR-ODS column (Shimadzu; column dimensions 3.0 mm $\times$ 75 mm, bead size 2.2 μm) and eluted at a flow rate of 0.7 ml min^{-1} at an increasing concentration of acetonitrile from 10% to 35% over 8.6 min in 0.1% aqueous formic acid. Sinapic acid, *p*-coumaric acid and kaempferol were used as standards for quantification of sinapate esters, *p*-coumarate esters and flavonols, respectively.

Cell wall saccharification assay. Pre-treatment was carried out by pressure cooking 50 mg samples in a metal tube containing 1.5 ml water at 200 °C (30 s of heat-up time followed by a 10 min hold). Each tube was placed in a fluidized sand bath (Tecam SBL-1; Cole-Parmer). The pressure within the tubes was held at the saturation vapour pressure of water to keep the water in a liquid state. The samples were cooled before the addition of 1.5 ml of 100 mM citrate buffer (pH 4.8), bringing the final volume to 3 ml (~2% solids (w/v)). The enzyme hydrolysis for all the conditions tested was based on initial solids loading and glucan concentration. Commercial cellulase (Spezyme CP) at 50 filter paper unit (FPU) g^{-1} glucan (90 mg protein per g glucan) and β-glucosidase (Novozyme 188) at 105 cellobiohydrolase unit (CBU) g^{-1} glucan (34 mg protein per g glucan) were added, and hydrolysis was carried out for different lengths of time at 50 °C and pH 4.8 in an incubator shaker (New Brunswick Scientific). The ratio of enzyme to solids was equivalent to 10 FPU g^{-1} total solids, 21 CBU g^{-1} total solids and 25 mg protein per g total solids. Enzyme hydrolysis of 50 mg untreated samples (also at ~2% solids (w/v), 50 °C and pH 4.8) was carried out under similar experimental conditions.

GPC. Acetylated CELs isolated from *Arabidopsis* stem cell walls were dissolved in dimethylformamide containing 0.1 M lithium bromide (~0.5 mg ml^{-1}), and subjected to GPC analysis on a Shimadzu LC-20A LC system equipped with SPD-M20A photodiode array (PDA) detector using the following conditions[38]: column, Tosoh TSK gel α-M plus α-2500; eluent, dimethylformamide with 0.1 M lithium bromide; flow rate, 0.5 ml min^{-1}; column oven temperature, 40 °C; sample detection, PDA response at 280 nm. The molecular weight calibration was via polystyrene standards. Data acquisition and computation was done using Shimadzu LCsolution v.1.25 software.

High-throughput mRNA sequencing. For global transcript analysis, we harvested total RNA from wild-type, *med5a/5b*, *ref8-1* and *med5a/5b ref8-1* plants. Three samples were collected for each genotype, with each sample consisting of three whole rosettes for wild-type, *med5a/5b* and *med5a/5b ref8-1* plants, and five whole rosettes for *ref8-1* plants. All plants of a given sample were grown in parallel in the same pot. Samples were harvested and flash frozen in liquid nitrogen on day 19 after planting, 6.5 h after subjective dawn, over a period of approximately 10 min. Frozen plant tissue was then ground in liquid nitrogen using a mortar and pestle, and total RNA was extracted using the Qiagen RNeasy Plant Mini Kit (Qiagen Sciences) according to the manufacturer's instructions. RNA samples were treated with DNase and purified using the Zymo Research DNA-Free RNA Kit (Zymo Research) according to the manufacturer's instructions.

cDNA libraries for sequencing were prepared using the Illumina TruSeq Kit (Illumina) according to the manufacturer's instructions. Libraries were then titrated using a KAPA quantification kit (Kapa Biosystems), denatured and clustered on both lanes of an Illumina Rapid Chemistry flow cell at 12 pM. Seventy-six bases of sequence were collected from each of two paired-end reads, as well as an index read to assign each pair of reads to its sample of origin.

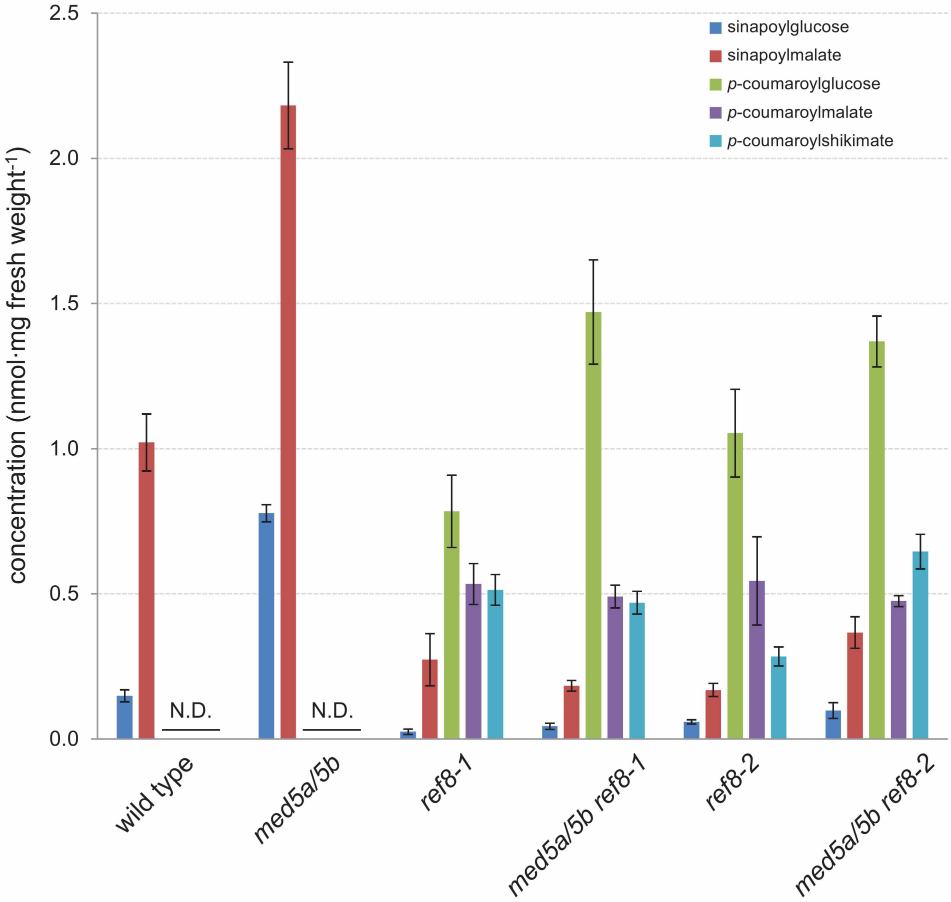
To pre-process filtered Illumina reads for mapping, sequences were first assessed for quality using FastQC (v.0.10.1; http://www.bioinformatics.babraham.ac.uk), then trimmed using the FASTX toolkit (v.0.0.13.2; http://hannonlab.cshl.edu) to remove bases with a Phred33 score less than 30. Trimmed reads less than 40 bases in length were discarded. Quality-trimmed reads were then mapped to the bowtie2-indexed *Arabidopsis* genome using Tophat (v.2.0.9; http://tophat.cbcb.umd.edu/) with default parameters. The total number of reads from all 12 samples that unambiguously mapped to a gene feature was 309,544,562 (average 25,795,380; minimum 17,571,930; maximum 30,794,071). A counts matrix consisting of the raw read count for each gene feature in each sample was generated using HTSeq (v.0.5.3p7; http://www-huber.embl.de/users/anders/HTSeq/). To eliminate fold changes of infinity and divide-by-zero errors, the counts matrix was modified such that genes with 0 counts across all samples were removed, and remaining 0 count values were changed to 1.

Differential expression analysis was performed using the statistical program R (v.3.0.1; http://www.r-project.org/) in conjunction with three analytical methods available from Bioconductor (http://www.bioconductor.org): DESeq (v.1.12.1), edgeR (v.3.2.4) and voom{limma} (v.3.16.6), as well as the Cufflinks (v.2.1.1) suite of programs (http://cufflinks.cbcb.umd.edu/). All four statistical methods gave similar overall conclusions. We selected the most conservative results (DESeq; false discovery rate < 0.05) for further investigation and reporting. Venn diagrams were adapted from those created with the online tool Venny (http://bioinfogp.cnb.csic.es/tools/venny/) and gene ontology term analysis was performed using the online tool DAVID (v.6.7; http://david.abcc.ncifcrf.gov/home.jsp).

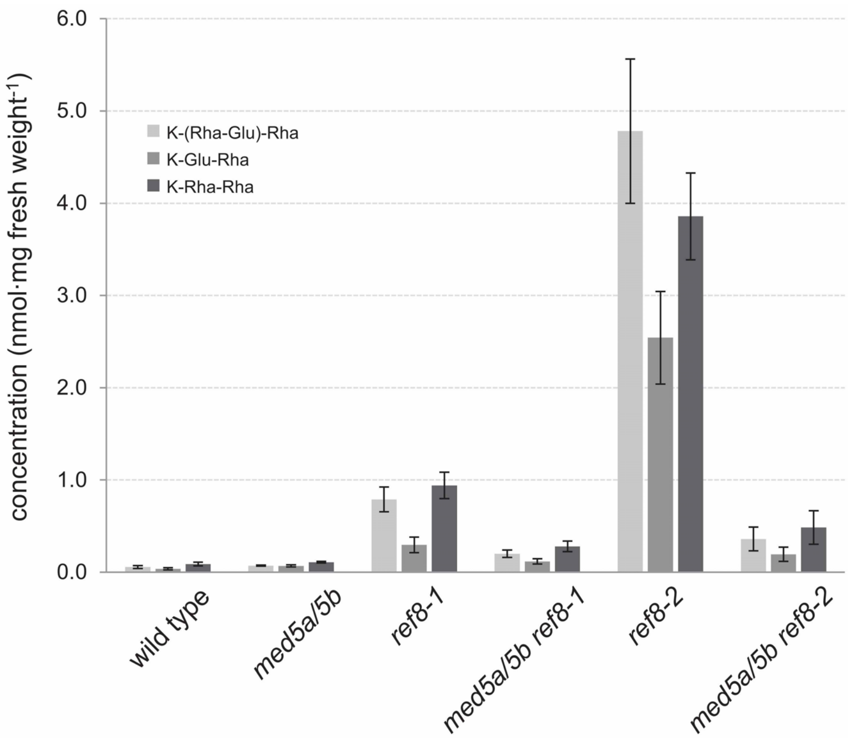
Quantification of salicylic acid. To extract salicylic acid from wild-type, *med5a/5b*, *ref8-1* and *med5a/5b ref8-1* plants, triplicate samples of 4-week-old, soil-grown plants were harvested, weighed and flash frozen in liquid nitrogen. Each sample consisted of a sufficient number of whole rosettes such that the total mass of the sample was at least 300 mg. Samples were ground in liquid nitrogen with a mortar and pestle and extracted at 4 °C for 24 h with mild shaking at a tissue concentration of 50 mg ml^{-1} in ice-cold methanol containing 500 μM 3,4,5-trimethoxy-*trans*-cinnamic acid as an internal standard. Water and chloroform were then added such that the ratio of methanol:water:chloroform was 2:1.2:1. Samples were then briefly vortexed and placed at 4 °C for an additional 12 h, after which the upper phase was transferred to a new tube, dried in a centrifugal evaporator (speedvac) and redissolved in 100 μl of 65% methanol (v/v) per 300 mg of tissue originally used for extraction.

Quantification of SA was accomplished by HPLC-MS/MS on an Agilent 1200 system using a Waters Xterra MS C18 column (5 μm, 150 mm $\times$ 2.1 mm inner diameter). A binary mobile phase consisting of 50 mM aqueous ammonium acetate pH 7.0 (solvent A) and acetonitrile (solvent B) was used at a flow rate of 0.3 ml min^{-1}. Initial conditions were set at 90:10 A:B with a linear gradient to 80:20 from 0 to 3 min, followed by a linear gradient to 30:70 from 3 to 13 min, followed by column re-equilibration. After separation, the column effluent was introduced by negative mode electrospray ionization (ESI) into an Agilent 6460 triple quadrupole mass spectrometer. ESI capillary voltage was -3.2 kV, nebulizer gas pressure was set at 35 psi, drying gas temperature was 325 °C, flow rate was 9 l min^{-1}, fragmentor voltage was set to 85 V. Multiple reaction monitoring transitions were 137.0 to 93.0 for SA and 237.2 to 103.1 for the internal standard. Collision energies used were 14 eV for SA and 10 eV for the internal standard. Mass data were collected and analysed using Agilent MassHunter software (v.B.02). SA quantification was accomplished using a standard curve over the range 0.01–50 μg ml^{-1}.

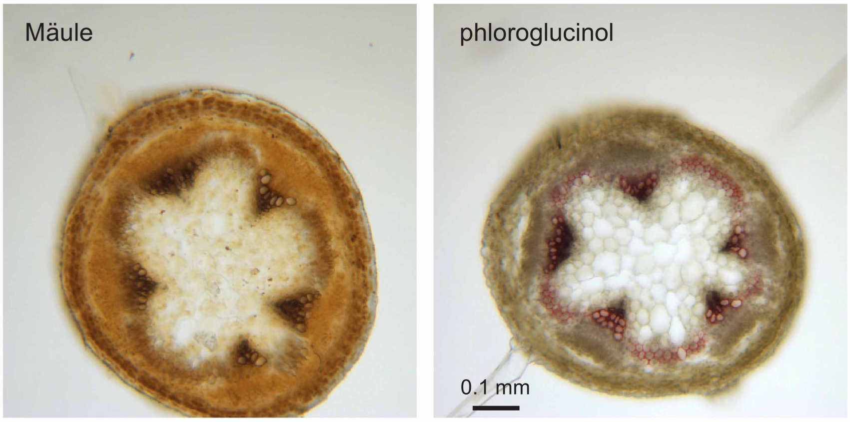
31. Stout, J., Romero-Severson, E., Ruegger, M. O. & Chapple, C. Semidominant mutations in reduced epidermal fluorescence 4 reduce phenylpropanoid content in *Arabidopsis*. *Genetics* **178,** 2237–2251 (2008).
32. Meyer, K. *et al.* Lignin monomer composition is determined by the expression of a cytochrome P450-dependent monooxygenase in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **95,** 6619–6623 (1998).
33. Wagner, A. *et al.* CCoAOMT suppression modifies lignin composition in *Pinus radiata*. *Plant J.* **67,** 119–129 (2011).
34. Weng, J. K., Akiyama, T., Ralph, J. & Chapple, C. Independent recruitment of an *O*-methyltransferase for syringyl lignin biosynthesis in *Selaginella moellendorffii*. *Plant Cell* **23,** 2708–2724 (2011).
35. Mansfield, S. D., Kim, H., Lu, F. & Ralph, J. Whole plant cell wall characterization using solution-state 2D NMR. *Nature Protoc.* **7,** 1579–1589 (2012).
36. Quideau, S. & Ralph, J. Facile large-scale synthesis of coniferyl, sinapyl, and *p*-coumaryl alcohol. *J. Agric. Food Chem.* **40,** 1108–1110 (1992).
37. Tobimatsu, Y. *et al.* Hydroxycinnamate conjugates as potential monolignol replacements: *in vitro* lignification and cell wall studies with rosmarinic acid. *ChemSusChem* **5,** 676–686 (2012).
38. Chen, F. *et al.* A polymer of caffeyl alcohol in plant seeds. *Proc. Natl Acad. Sci. USA* **109,** 1772–1777 (2012).



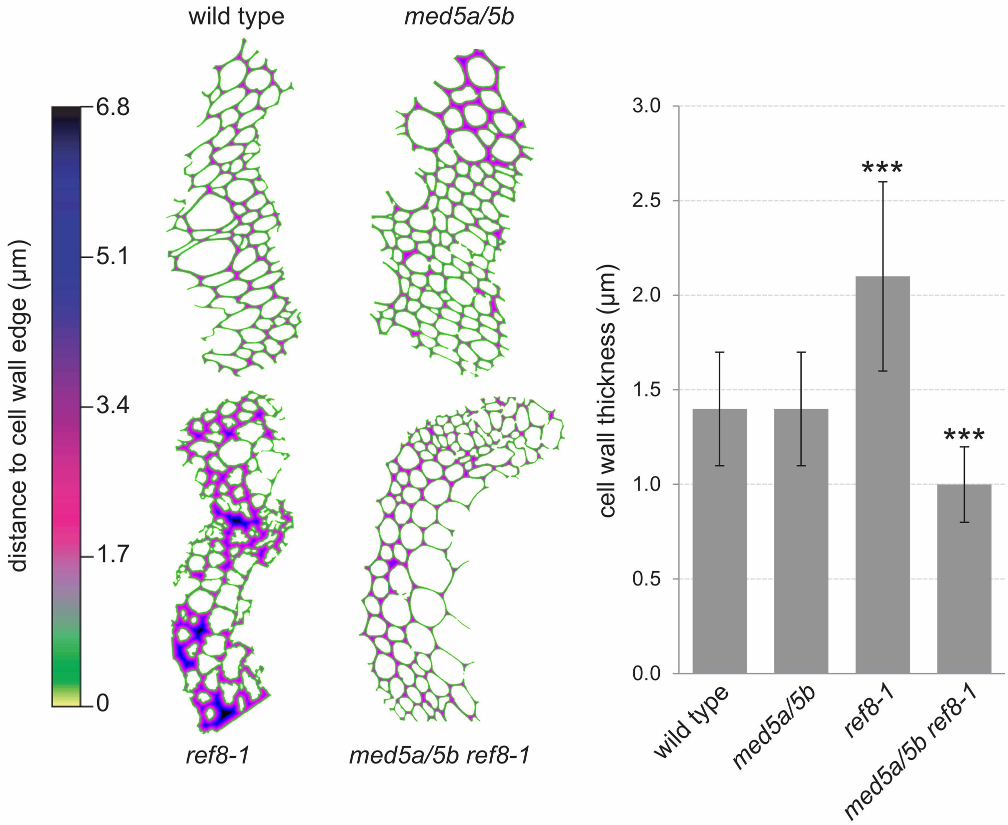
Extended Data Figure 1 | Disruption of *MED5a/5b* does not restore normal hydroxycinnamate ester biosynthesis to *ref8* mutants. Quantification of the hydroxycinnamate esters sinapoylglucose, sinapoylmalate, *p*-coumaroylglucose, *p*-coumaroylmalate and *p*-coumaroylshikimate in rosettes of 3-week-old plants of the indicated genotypes. Similar to *ref8* mutants, *med5a/5b ref8* plants fail to accumulate wild-type levels of the sinapate esters sinapoylglucose and sinapoylmalate, and instead hyperaccumulate the *p*-coumarate esters *p*-coumaroylglucose, *p*-coumaroylmalate and *p*-coumaroylshikimate. In each case, data are derived from five individual plants, except for *ref8-2*, where owing to their small size, three groups of five, seven and seven plants each were combined into independent pools. Error bars indicate s.d. in all cases.



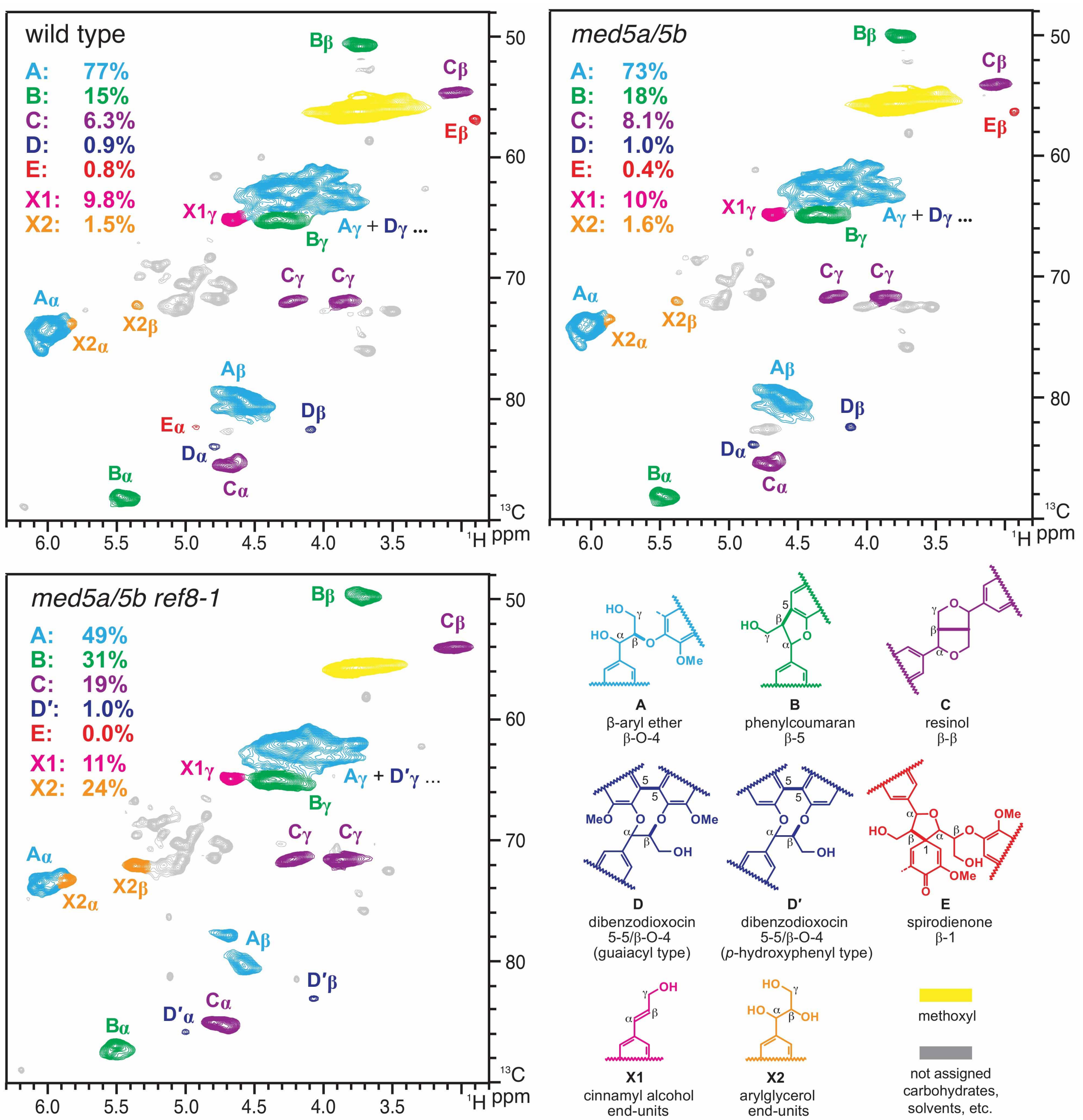
Extended Data Figure 2 | Disruption of *MED5a/5b* alleviates the flavonoid hyperaccumulation of *ref8* mutants. Quantification of the three major flavonol glucosides in rosettes of 3-week-old plants of the indicated genotypes. In each case, data are derived from five single plants, except for *ref8-2*, where three groups of five, seven and seven plants each were combined into independent pools. Error bars indicate s.d. K-(Rha-Glu)-Rha, kaempferol 3-*O*-[6′′-*O*-(rhamnosyl) glucoside] 7-*O*-rhamnoside; K-Glu-Rha, kaempferol 3-*O*-glucoside 7-*O*-rhamnoside; K-Rha-Rha, kaempferol 3-*O*-rhamnoside 7-*O*-rhamnoside.



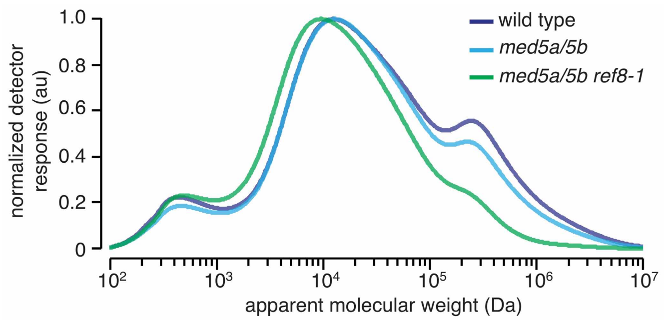
Extended Data Figure 3 | *med5a/5b ref8-2* mutants show patterns of lignification similar to *med5a/5b ref8-1* mutants. Shown are thin sections of inflorescence stems of *med5a/5b ref8-2* mutants stained with Mäule reagent (left) and phloroglucinol (right). Plants were grown and stained in parallel with those shown in Fig. 2a, b. Although *med5a/5b ref8-2* mutant stems are somewhat thinner than those of *med5a/5b ref8-1* mutants and show some morphological abnormalities, the overall staining patterns of *med5a/5b ref8-1* and *med5a/5b ref8-2* inflorescence stems are highly similar. The corresponding tissues of *ref8-2* mutant plants could not be examined owing to their developmental arrest shortly after germination.



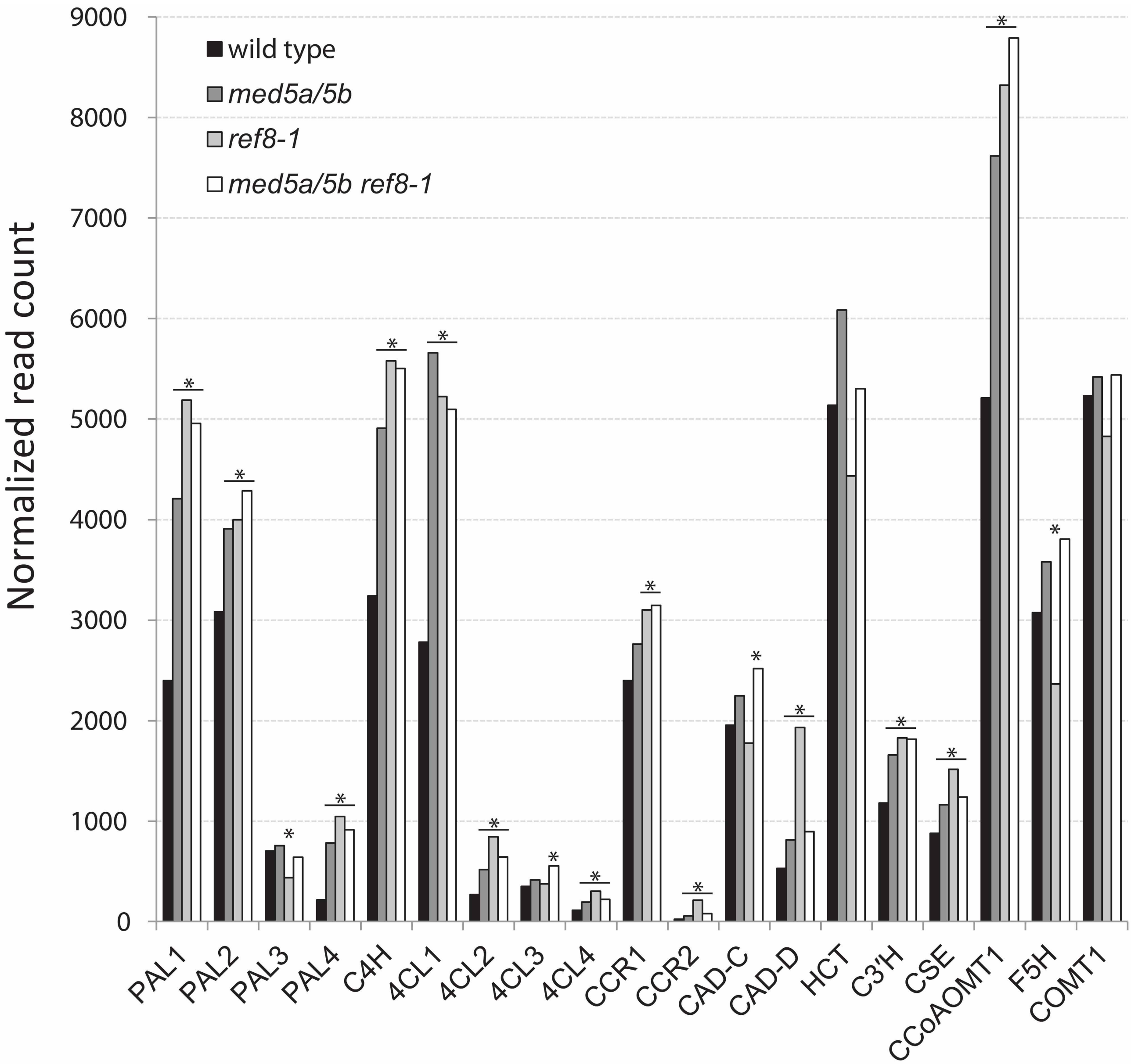
Extended Data Figure 4 | *ref8-1* mutants show thickening of the secondary cell wall that is rescued by disruption of *MED5a/5b*. Left, distance map of cell wall thickness calculated from micrographs of representative samples of stem cross-sections of plants of the indicated genotypes. Right, quantification of cell wall thickness. $N > 200$ cells, with at least 100 measurements per cell for each sample. Error bars represent s.d. ***$P < 0.001$, difference from wild type (Student's t-test).



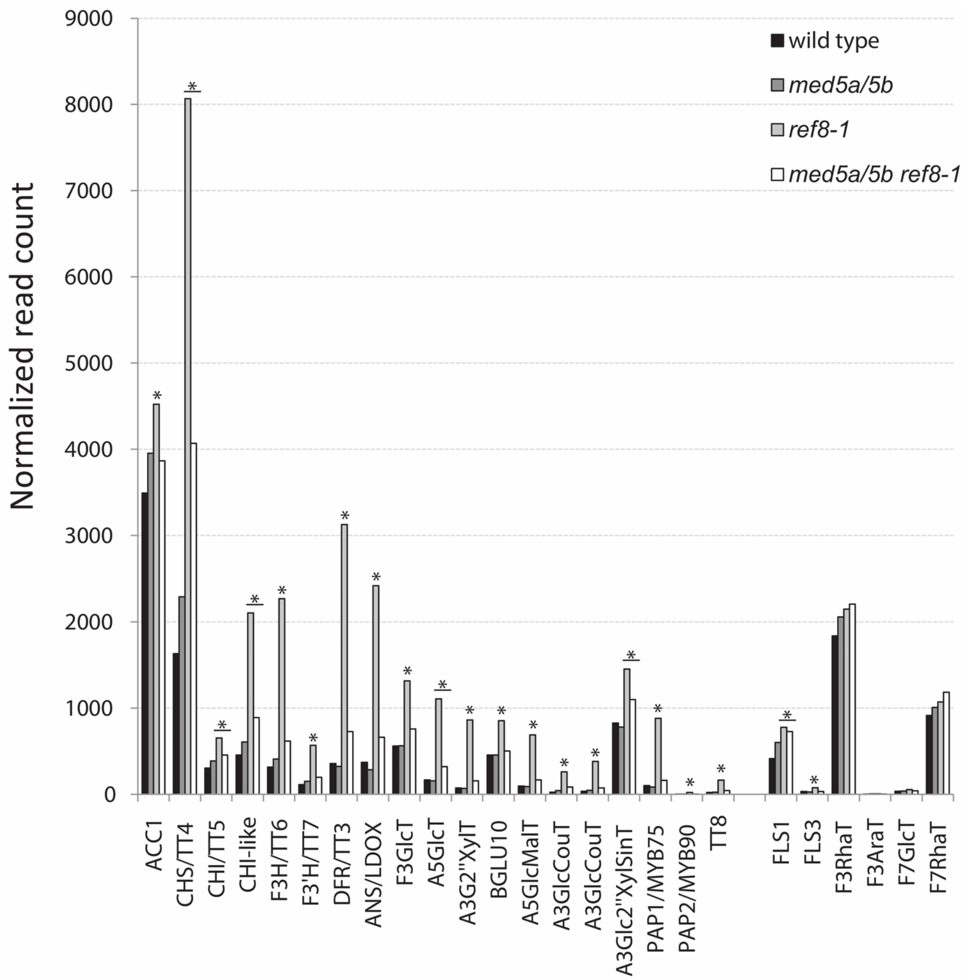
Extended Data Figure 5 | Lignin of *med5a/5b ref8-1* mutant plants differs structurally from lignin of wild-type or *med5a/5b* mutant plants. 2D-NMR spectra of lignin from the indicated genotypes. The data shown are derived from a different region of the same spectra shown in Fig. 3. Colour-coded structures on right correspond to the major resonances in each spectrum.



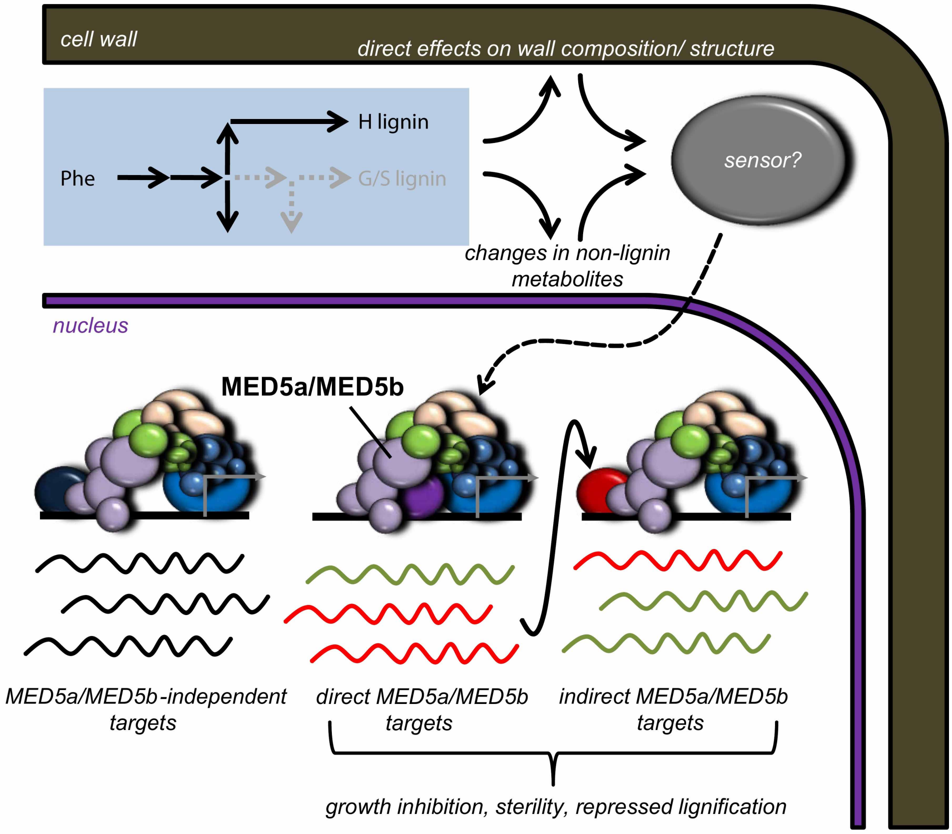
Extended Data Figure 6 | High-molecular-weight lignin polymers are underrepresented in *med5a/5b ref8-1* mutants. Shown are the results of gel-permeation chromatography of lignin from wild-type, *med5a/5b* and *med5a/5b ref8-1* cell walls. The *x*-axis indicates the apparent molecular weight of individual lignin polymer fragments and is shown as a log scale. The *y*-axis shows the response of an ultraviolet-light detector normalized to the most abundant signal in each chromatogram. The most abundant signal in all samples corresponds to a molecular weight of ~10,000 Da, whereas a secondary peak at ~250,000 Da is significantly underrepresented in lignin derived from the *med5a/5b ref8-1* mutant.



Extended Data Figure 7 | Expression of lignin biosynthesis genes in wild-type, *med5a/5b*, *ref8-1* and *med5a/5b ref8-1* plants. Shown is the expression of general phenylpropanoid and lignin biosynthesis genes in 3-week-old rosettes of plants of the indicated genotypes as determined using high-throughput sequencing of mRNA. The value shown on the *y*-axis refers to the number of reads unambiguously mapping to each gene, normalized for differences in the total number of reads between samples and for lane effects. *$P < 0.05$, difference from wild type, as determined by the DESeq algorithm using a Benjamini–Hochberg procedure to adjust for multiple testing.

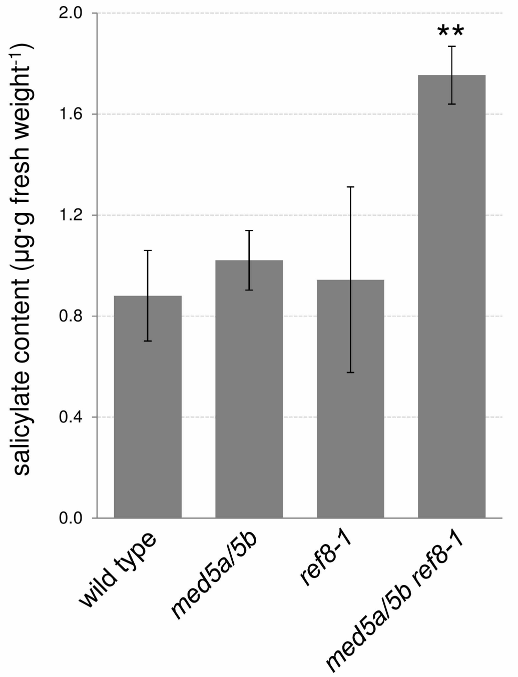


Extended Data Figure 8 | Expression of flavonoid biosynthesis genes in wild-type, *med5a/5b*, *ref8-1* and *med5a/5b ref8-1* plants. Shown is the expression of flavonoid biosynthesis genes in 3-week-old rosettes of plants of the indicated genotypes as determined using high-throughput sequencing of mRNA. The value shown on the *y*-axis refers to the number of reads unambiguously mapping to each gene, normalized for differences in the total number of reads between samples and for lane effects. $*P < 0.05$, difference from wild type, as determined by the DESeq algorithm using a Benjamini–Hochberg procedure to adjust for multiple testing.



Extended Data Figure 9 | A model for Mediator-dependent growth inhibition in *Arabidopsis ref8* mutants. Mutation or disruption of *REF8* leads to direct alterations in the composition of the cell wall and other metabolic changes due to the loss of C3′H activity. Information on these changes is relayed to the nucleus by an at present unknown signalling pathway or sensor, resulting in massive changes in gene expression (represented by green and red transcripts in the model). Some of these transcriptional changes are directly dependent on MED5 (centre, illustrated as a direct MED5–transcription-factor interaction), whereas others are independent of MED5 (left) or are indirectly affected by MED5 (right), such as genes controlled by transcription factors that are themselves MED5-dependent targets. Ultimately, changes in the transcription of direct and/or indirect targets of MED5 result in inhibition of growth, sterility and indirect effects on cell wall architecture, all of which can be rescued by disruption of MED5.

a



b

Extended Data Figure 10 | *med5a/5b ref8-1* mutants show elevated levels of salicylic acid and disruption of SID2 does not rescue the stunted growth of the *ref8-1* mutant. a, Shown is the quantification of salicylic acid in 3-week-old rosettes of plants of the indicated genotypes. Data for each genotype are derived from three independently pooled samples representing 300 mg of whole rosette tissue each. Error bars represent s.d. $^{**}P < 0.01$, difference from wild type (Student's *t*-test). **b**, Shown are 3-week-old rosettes of representative plants of the indicated genotypes. The *SID2* gene encodes the salicylic acid biosynthetic enzyme isochorismate synthase. The *sid2-4* and *ref8-1 sid2-4* plants shown are representative progeny of a single plant with the genotype *sid2-4*/*sid2-4 REF8-1*/*ref8-1* that gave rise to both morphologically normal and dwarfed plants at a ratio of 3:1. $N = 167$ morphologically normal, 50 dwarfed; $\chi^2 = 0.444$, $P = 0.502$.

LETTER

doi:10.1038/nature13117

Structure of the core ectodomain of the hepatitis C virus envelope glycoprotein 2

Abdul Ghafoor Khan[1], Jillian Whidby[1], Matthew T. Miller[1], Hannah Scarborough[2], Alexandra V. Zatorski[1], Alicja Cygan[1], Aryn A. Price[2], Samantha A. Yost[1], Caitlin D. Bohannon[2], Joshy Jacob[2], Arash Grakoui[2,3] & Joseph Marcotrigiano[1]

Hepatitis C virus (HCV) is a significant public health concern with approximately 160 million people infected worldwide[1]. HCV infection often results in chronic hepatitis, liver cirrhosis and hepatocellular carcinoma. No vaccine is available and current therapies are effective against some, but not all, genotypes. HCV is an enveloped virus with two surface glycoproteins (E1 and E2). E2 binds to the host cell through interactions with scavenger receptor class B type I (SR-BI) and CD81, and serves as a target for neutralizing antibodies[2–4]. Little is known about the molecular mechanism that mediates cell entry and membrane fusion, although E2 is predicted to be a class II viral fusion protein. Here we describe the structure of the E2 core domain in complex with an antigen-binding fragment (Fab) at 2.4 Å resolution. The E2 core has a compact, globular domain structure, consisting mostly of β-strands and random coil with two small α-helices. The strands are arranged in two, perpendicular sheets (A and B), which are held together by an extensive hydrophobic core and disulphide bonds. Sheet A has an IgG-like fold that is commonly found in viral and cellular proteins, whereas sheet B represents a novel fold. Solution-based studies demonstrate that the full-length E2 ectodomain has a similar globular architecture and does not undergo significant conformational or oligomeric rearrangements on exposure to low pH. Thus, the IgG-like fold is the only feature that E2 shares with class II membrane fusion proteins. These results provide unprecedented insights into HCV entry and will assist in developing an HCV vaccine and new inhibitors.

HCV envelope glycoprotein 2 (E2) is a type I transmembrane protein with an amino-terminal ectodomain connected to a carboxy-terminal transmembrane helix through an amphipathic, α-helical stem (Fig. 1a)[5,6]. E2 is highly modified post-translationally with 9–11 *N*-linked glycosylation sites and 18 cysteine residues that are conserved across all genotypes. For ease of comparison with other genotypes, we refer to the cysteines and *N*-linked glycosylation sites as C1 to C18 and N1 to N11, respectively, with residue numbers from the J6 genotype (2a) given in parentheses. Full-length, E2 ectodomain (eE2) (384–656) was produced in *N*-acetylglucosaminyltransferase I-negative (GnTI−) HEK293T cells by a lentiviral expression system and grown in an adherent cell bioreactor. The resulting eE2 protein is monomeric as determined by non-reducing SDS–polyacrylamide gel electrophoresis (PAGE) and size-exclusion chromatography (Extended Data Fig. 1).

Solution-based studies using limited proteolysis and hydrogen deuterium exchange demonstrated that approximately 80 amino acids on the N terminus (384–463) from hypervariable region (HVR) 1 through to HVR2 are exposed and flexible. This region includes conserved sequences implicated in binding to the cellular receptors (SR-BI and CD81) as well as several epitopes for neutralizing antibodies (Fig. 1 and Extended Data Figs 2 and 3)[7–11]. Various N-terminal deletions were produced to minimize regions of disorder while preserving an even number of cysteines, potentially allowing for intramolecular disulphide-bond formation. All constructs were screened for aggregation by non-reducing SDS–PAGE and size-exclusion chromatography. E2 core (456–656) is soluble, monomeric and maintains similar secondary structure content when compared with eE2 as determined by reactivity towards HCV-infected patient sera (Extended Data Fig. 4a, b) and circular dichroism (data not shown). However, in contrast to eE2, CD81 binding affinity and the efficiency of inhibition of HCV cell culture (HCVcc) entry was diminished for the E2 core (Extended Data Fig. 4c–e). This indicates that the N terminus of eE2 is critical for CD81 interaction and probably undergoes a transition from disorder to order on binding. Alternatively, the N-terminal region may also be ordered through interactions with other factors, for example, E1, apolipoproteins, lipids, cellular receptors, or antibodies.

Monoclonal antibodies were generated against recombinant eE2 and crystals of deglycosylated E2 core were produced in complex with a Fab (2A12) and diffracted to 2.4 Å resolution (Fig. 1 and Extended Data Table 1). The complex structure was determined by molecular replacement using a Fab structure followed by iterative rounds of model building and refinement. The E2 core domain has a globular fold, consisting of mostly β-strands and random coil with two short α-helices, which is consistent with previous spectroscopic studies of eE2 (refs 12, 13). The protein contains two, four-stranded antiparallel β-sheets (termed sheets A and B), the planes of which are approximately perpendicular to each other. The four strands of the N-terminal β-sheet (sheet A) are stabilized by two disulphide bonds, between strands 1 and 3 (C7 (510) and C8 (554)) and the N-terminal loop with strand 4 (C5 (496) and C9 (566)). The loop between strands 2 and 3 contains sequences implicated in CD81 binding and is flexible, similar to the N-terminal CD81 binding sites, which were deleted[14,15]. After strand 4, the polypeptide continues into a long, disordered loop before forming the first short helix (H1) followed by the second β-sheet (sheet B). A second short α-helix (H2) is located between strands 6 and 7. A disulphide bond (C14 (611) and C16 (648)) between strand 6 and the C-terminal strand 8 further stabilizes the fold. The C-terminal strands (7 and 8) are the longest within the protein with approximately nine amino acids each and encompass the 2A12-binding site. 2A12 does not neutralize HCV infection, indicating that the epitope is either buried within the particle or incapable of preventing entry (Extended Data Fig. 4f). The two β-sheets are held together by (1) two disulphide bonds, connecting the loops before strand 1 and after H2 (C4 (488) with C15 (624)) as well as the loops after strand 4 and before H1 (C10 (571) and C13 (601)), and (2) an extensive hydrophobic core consisting of numerous aromatic residues (Extended Data Fig. 5).

HCV belongs to the genus *Hepacivirus* of the Flaviviridae family. Other members of the family include the flavivirus and pestivirus genera, which consist of arthropod-borne viruses and important livestock pathogens, respectively[16]. The flavivirus envelope glycoprotein (E) is a class II fusion protein and HCV E2 was expected to have a similar fold[13,17,18]. All class II fusion proteins have a common elongated structure, consisting of predominantly β-sheets, and exist as homo- or

[1]Center for Advanced Biotechnology and Medicine, Department of Chemistry and Chemical Biology, Rutgers University, 679 Hoes Lane West, Piscataway, New Jersey 08854, USA. [2]Division of Microbiology and Immunology, Emory Vaccine Center, Emory University School of Medicine, 100 Woodruff Circle, Atlanta, Georgia 30322, USA. [3]Division of Infectious Diseases, Department of Medicine, Emory University School of Medicine, 100 Woodruff Circle, Atlanta, Georgia 30322, USA.

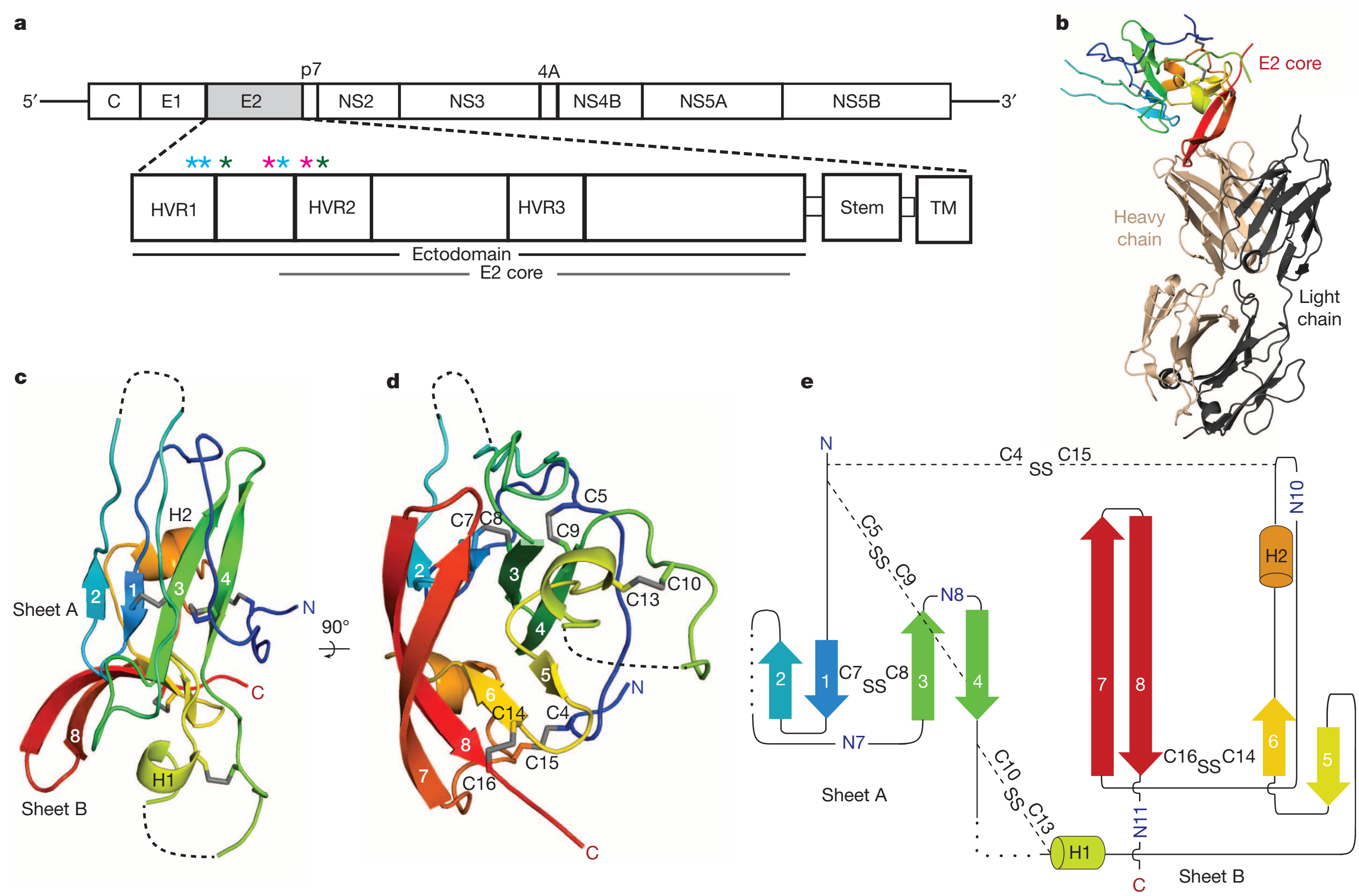


Figure 1 | Overview of HCV E2. **a**, Schematic representation of the HCV genome and E2 domain organization. Full-length eE2 and the crystallization construct are indicated by the black and grey bars, respectively. C, capsid protein; non-structural protein (NS) 2-5B. Asterisks indicate the location of trypsin (blue), chymotrypsin (green) and GluC (magenta) cleavage sites. **b**–**d**, Ribbon diagram of the E2 core domain bound to Fab 2A12 (**b**) and alone (**c** and **d**). The view in **d** is a 90° rotation about a horizontal axis from **c**. The E2 polypeptide chain is coloured from the N terminus (blue) to C terminus (red). **e**, Topology diagram of E2 core domain, detailing secondary structure elements, disulphide bonds (dashed lines labelled with SS), *N*-linked glycosylation sites and regions of disordered polypeptide (dotted lines).

heterodimers with the membrane-fusion, hydrophobic peptide buried at the dimer interface at neutral pH. On receptor binding and/or exposure to low pH, these proteins undergo self-rearrangement into stable trimers, exposing the fusion peptide and resulting in viral and host membrane fusion. Despite containing a similar extended organization, the recent structure of the pestivirus bovine viral diarrhoea virus (BVDV) E2 glycoprotein does not represent a typical class II fusion protein fold and lacks an apparent fusion peptide, indicating that it is unlikely to be a class II fusion protein[19,20].

Similar to the flavivirus and pestivirus glycoproteins, the HCV E2 core secondary structure consists of predominantly β-sheets and random coil. However, E2 core is a monomer with a compact globular shape, in contrast to the extended structures reported in other viruses. Solution-based small-angle X-ray scattering (SAXS) was used to correlate the crystallographic core domain structure with fully glycosylated eE2 and various fragments. The *ab initio* SAXS envelopes of E2 core and eE2 are similar, with approximately the same radius of gyration (R_g) (Fig. 2a, b and Extended Data Table 2). Glycosylation, which is missing in the E2 core crystal structure, represents roughly one-third of the mass and accounts for the unmodelled areas of the envelopes. Notably, neither the R_g nor the elution profiles on size-exclusion chromatography for fully glycosylated eE2 and E2 core changed significantly at pH 5.0 (Extended Data Fig. 6a, b). These results indicate that unlike class II membrane fusion proteins, E2 does not undergo significant structural rearrangements on exposure to low pH.

SAXS was used to investigate the CD81 binding region on the E2 ectodomain. To simplify data interpretation, eE2(ΔHVR1) was used, as HCV lacking HVR1 remains infectious[21]. The binding site of CD81 was identified by superimposing the SAXS envelopes of eE2(ΔHVR1) alone and in complex with CD81-LEL (Fig. 2c–e). Although CD81-LEL is a dimer in solution (Extended Data Fig. 6c), the extra density in the SAXS envelope is more consistent with monomeric binding; however, a dimer cannot be ruled out.

HCV E2 is modified by *N*-linked glycosylation, which is necessary for proper folding and immune invasion. E2 from the J6 genotype has 11 glycosylation sites. Four of the glycosylation sites are in the flexible N-terminal region, which were deleted, and seven are in the core domain (N5–N11). The location of N7, N8, N10 and N11 are modelled in the final E2 core structure. All of these glycans are present in loop areas, indicating that these sites are solvent exposed and flexible. Mutagenesis studies in HCVcc have shown that N6, N8 and N10 are integral for virus infectivity. Removal of the N6 site results in improved CD81 binding, whereas N8 and N10 mutations destabilize the protein and cause defective particle production[22]. Both sheets have one critical glycosylation site: N8 in sheet A and N10 in sheet B. All four of the observed glycosylation sites are on the periphery of the core and are located on a highly basic surface (Fig. 3). The opposite surface is predominantly hydrophobic and highly conserved when compared to the basic surface. Furthermore, the epitope for antibodies (AR1, AR3A and AR3C) that inhibit E1E2 binding to CD81 is located at the interface of

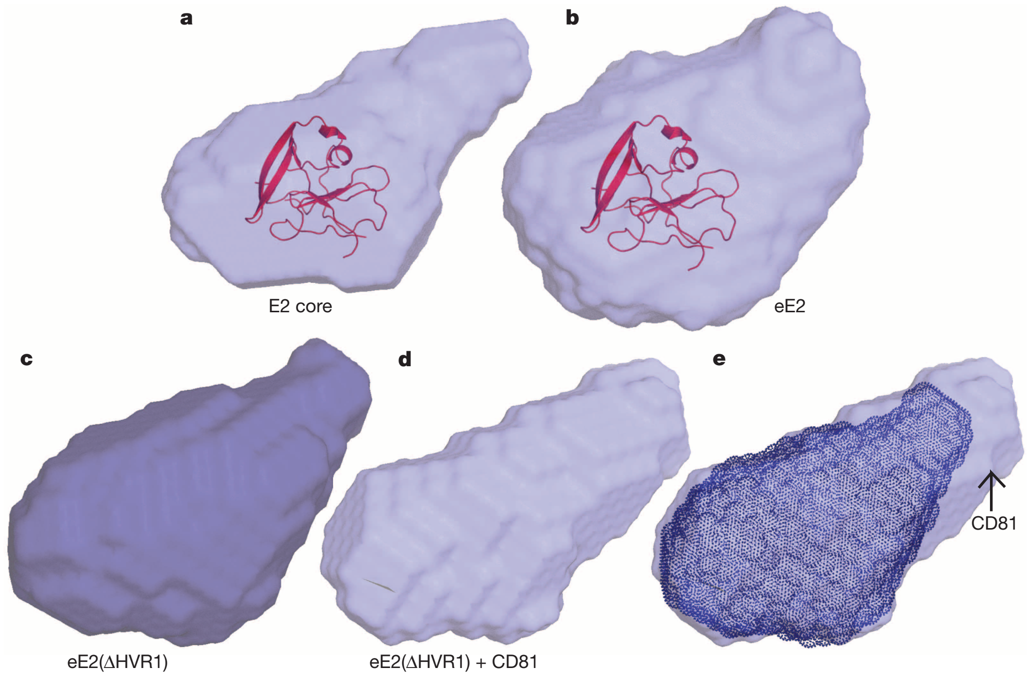


Figure 2 | *Ab initio* SAXS envelopes of E2 core, eE2(ΔHVR1) and eE2. **a**–**d**, SAXS envelopes of glycosylated E2 core (**a**), eE2 (**b**), eE2(ΔHVR1) (**c**) and eE2(ΔHVR1) in complex with CD81-LEL (**d**). The E2 core domain structure has been fitted into **a** and **b**. **e**, Superposition of the SAXS envelopes of eE2(ΔHVR1) alone (**c**) and in complex with CD81-LEL (**d**), highlighting the approximate position of CD81-LEL.

the hydrophobic and basic surfaces, including the N7 glycosylation site (Extended Data Fig. 7a). Interestingly, N7 is only 7 residues away from N6, which has a critical effect on CD81 binding. Epitopes for antibodies (that is, AR5) that block E1E2 heterodimerization are also found on the hydrophobic surface, making it highly plausible that this surface is interacting with E1 in the context of the viral particle[4].

The precise roles played by E1 and E2 in membrane fusion are not fully understood. It has been predicted that amino acids 262–290 in E1

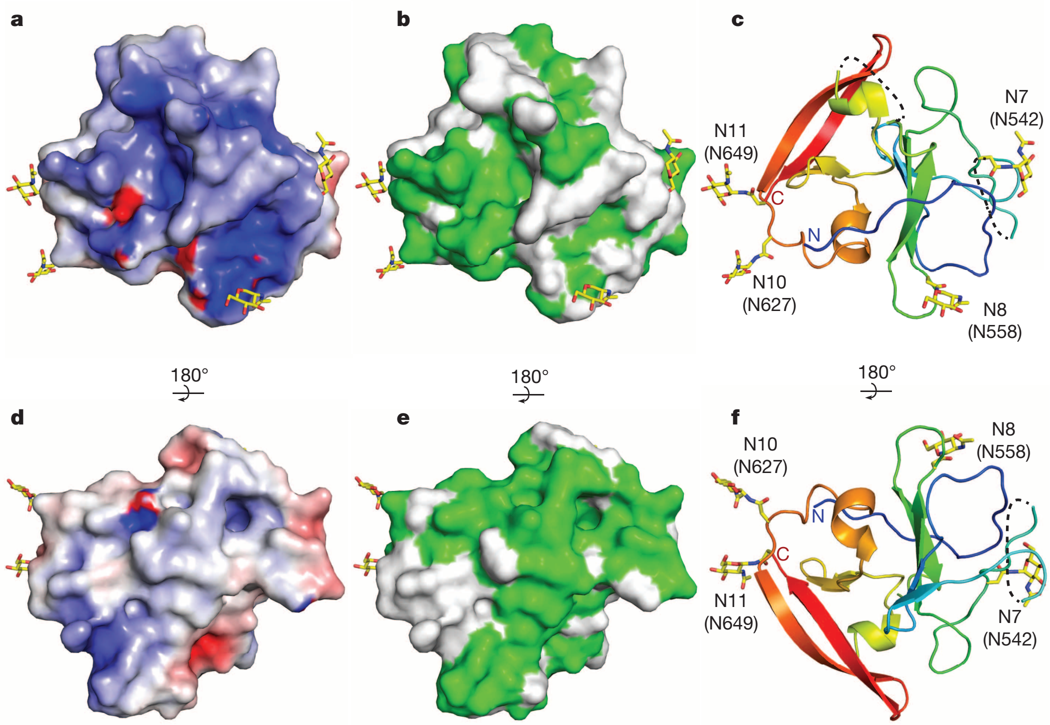


Figure 3 | Surface features of E2. The surface of the E2 core domain coloured for electrostatic potential (**a** and **d**)—blue (basic), white (neutral), red (acidic) at $\pm 5\,\mathrm{kT\,e^{-1}}$—and sequence identity (**b** and **e**) (green) from the alignment in Extended Data Fig. 2. **c**, **f**, Ribbon diagrams highlighting the location of the *N*-linked glycosylation sites. The orientations of **d**–**f** as well as **a**–**c** are identical. The orientation in **d**–**f** is rotated 180° about a horizontal axis from the view in **a**–**c**.

as well as 416–430, 504–522 and 604–624 in E2 are important for fusion[13,23,24]. In the structure, the potential fusion regions in E2 (504–522 and 604–624) are located in secondary structure elements within the hydrophobic core and therefore unlikely to serve as the fusion peptide. Furthermore, size-exclusion chromatography and SAXS analyses at low pH indicate that E2 does not undergo oligomeric or structural rearrangement. Thus, it seems unlikely that E2 has a direct role in membrane fusion. However, it is possible that E1 alone or the E1E2 heterodimer has a major role in the fusion process.

Structural comparison of the HCV E2 core domain with all known folds in the Protein Data Bank using the Dali server[25] identified proteins with IgG-like folds similar to the N-terminal sheet A, none of which is a class II fusion protein, although IgG-like folds are common in these proteins. The server failed to identify any statistically significant structures to sheet B, suggesting a novel fold. During the review process of this manuscript, a structure for HCV E2 from genotype 1a was published[26]. The core domain of both structures is highly similar with a root mean squared deviation of 0.8 Å for similar carbon α atoms. Our biochemical and structural data provide valuable information towards defining the role of E2 and establish a foundation for further studies in understanding HCV entry and infection.

METHODS SUMMARY

Recombinant, full-length E2 ectodomain (384–656, genotype J6), eE2(ΔHVR1) (413–656) and E2 core (456–656) were produced in HEK293T GnTI− cells[27] by a lentiviral expression system with a C-terminal, protein-A tag. Stable cell lines were grown in an adherent cell bioreactor. Collected supernatants were purified over IgG sepharose and the E2 proteins were eluted by PreScission Protease digestion. The eluted protein was deglycosylated with endoglycosidase H (EndoH) and purified to homogeneity by heparin affinity and size-exclusion chromatography. Monoclonal antibody 2A12 secreted by hybridoma cells maintained in suspension culture was purified over protein G resin. Fab was generated using papain digestion and subtractive protein A chromatography. E2 and 2A12 Fab complex was purified by size-exclusion chromatography and crystals were obtained in 18% (w/v) PEG 3350, 0.5 M $MgCl_2$, 0.1 M HEPES (pH 7.5), 15% dioxane and 2% formamide at 20 °C by the hanging-drop vapour diffusion method. The crystals belong to space group $P2_12_12$ with cell parameters $a = 85.96$ Å, $b = 194.57$ Å and $c = 37.92$ Å. The structure was determined by molecular replacement to 2.4 Å resolution using a mouse Fab (Protein Data Bank accession 2GSI). The final model has an R_{work} and R_{free} of 0.217 and 0.269, respectively. SAXS measurements were performed on eE2, E2 core, eE2(ΔHVR1) and eE2(ΔHVR1)/CD81-LEL. Data were analysed using BioXTAS RAW[28] and applications from the ATSAS[29] program suite.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 22 November 2013; accepted 31 January 2014.
Published online 19 February 2014.

1. Lavanchy, D. Evolving epidemiology of hepatitis C virus. *Clin. Microbiol. Infect.* **17,** 107–115 (2011).
2. Pileri, P. *et al.* Binding of hepatitis C virus to CD81. *Science* **282,** 938–941 (1998).
3. Scarselli, E. *et al.* The human scavenger receptor class B type I is a novel candidate receptor for the hepatitis C virus. *EMBO J.* **21,** 5017–5025 (2002).
4. Sautto, G., Tarr, A. W., Mancini, N. & Clementi, M. Structural and antigenic definition of hepatitis C virus E2 glycoprotein epitopes targeted by monoclonal antibodies. *Clin. Dev. Immunol.* **2013,** 450963 (2013).
5. Michalak, J. P. *et al.* Characterization of truncated forms of hepatitis C virus glycoproteins. *J. Gen. Virol.* **78,** 2299–2306 (1997).
6. Drummer, H. E. & Poumbourios, P. Hepatitis C virus glycoprotein E2 contains a membrane-proximal heptad repeat sequence that is essential for E1E2 glycoprotein heterodimerization and viral entry. *J. Biol. Chem.* **279,** 30066–30072 (2004).
7. Wahid, A. & Dubuisson, J. Virus-neutralizing antibodies to hepatitis C virus. *J. Viral Hepat.* **20,** 369–376 (2013).
8. Keck, Z. Y. *et al.* Human monoclonal antibodies to a novel cluster of conformational epitopes on HCV E2 with resistance to neutralization escape in a genotype 2a isolate. *PLoS Pathog.* **8,** e1002653 (2012).
9. Kong, L. *et al.* Structure of hepatitis C virus envelope glycoprotein E2 antigenic site 412 to 423 in complex with antibody AP33. *J. Virol.* **86,** 13085–13088 (2012).
10. Kong, L. *et al.* Structural basis of hepatitis C virus neutralization by broadly neutralizing antibody HCV1. *Proc. Natl Acad. Sci. USA* **109,** 9499–9504 (2012).
11. Deng, L. *et al.* Structural evidence for a bifurcated mode of action in the antibody-mediated neutralization of hepatitis C virus. *Proc. Natl Acad. Sci. USA* **110,** 7418–7422 (2013).
12. Whidby, J. *et al.* Blocking hepatitis C virus infection with recombinant form of envelope protein 2 ectodomain. *J. Virol.* **83,** 11078–11089 (2009).
13. Krey, T. *et al.* The disulfide bonds in glycoprotein E2 of hepatitis C virus reveal the tertiary organization of the molecule. *PLoS Pathog.* **6,** e1000762 (2010).
14. Keck, Z. Y. *et al.* Analysis of a highly flexible conformational immunogenic domain A in hepatitis C virus E2. *J. Virol.* **79,** 13199–13208 (2005).
15. Rothwangl, K. B., Manicassamy, B., Uprichard, S. L. & Rong, L. Dissecting the role of putative CD81 binding regions of E2 in mediating HCV entry: putative CD81 binding region 1 is not involved in CD81 binding. *Virol. J.* **5,** 46 (2008).
16. Lindenbach, B. D., Thiel, H.-J. & Rice, C. M. in *Fields Virology* (eds Knipe, D. M. & Howley, P. M.) 1101–1152 (Lippincott Williams & Wilkins, 2007).
17. White, J. M., Delos, S. E., Brecher, M. & Schornberg, K. Structures and mechanisms of viral membrane fusion proteins: multiple variations on a common theme. *Crit. Rev. Biochem. Mol. Biol.* **43,** 189–219 (2008).
18. Vaney, M. C. & Rey, F. A. Class II enveloped viruses. *Cell. Microbiol.* **13,** 1451–1459 (2011).
19. El Omari, K., Iourin, O., Harlos, K., Grimes, J. M. & Stuart, D. I. Structure of a pestivirus envelope glycoprotein E2 clarifies its role in cell entry. *Cell Rep.* **3,** 30–35 (2013).
20. Li, Y., Wang, J., Kanai, R. & Modis, Y. Crystal structure of glycoprotein E2 from bovine viral diarrhea virus. *Proc. Natl Acad. Sci. USA* **110,** 6805–6810 (2013).
21. Forns, X. *et al.* Hepatitis C virus lacking the hypervariable region 1 of the second envelope protein is infectious and causes acute resolving or persistent infection in chimpanzees. *Proc. Natl Acad. Sci. USA* **97,** 13318–13323 (2000).
22. Helle, F. *et al.* Role of N-linked glycans in the functions of hepatitis C virus envelope proteins incorporated into infectious virions. *J. Virol.* **84,** 11905–11915 (2010).
23. Lavillette, D. *et al.* Characterization of fusion determinants points to the involvement of three discrete regions of both E1 and E2 glycoproteins in the membrane fusion process of hepatitis C virus. *J. Virol.* **81,** 8752–8765 (2007).
24. Li, H. F., Huang, C. H., Ai, L. S., Chuang, C. K. & Chen, S. S. Mutagenesis of the fusion peptide-like domain of hepatitis C virus E1 glycoprotein: involvement in cell fusion and virus entry. *J. Biomed. Sci.* **16,** 89 (2009).
25. Holm, L. & Rosenstrom, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38,** W545–W549 (2010).
26. Kong, L. *et al.* Hepatitis C virus E2 envelope glycoprotein core structure. *Science* **342,** 1090–1094 (2013).
27. Reeves, P. J., Callewaert, N., Contreras, R. & Khorana, H. G. Structure and function in rhodopsin: high-level expression of rhodopsin with restricted and homogeneous N-glycosylation by a tetracycline-inducible N-acetylglucosaminyltransferase I-negative HEK293S stable mammalian cell line. *Proc. Natl Acad. Sci. USA* **99,** 13419–13424 (2002).
28. Nielsen, S. S., Moller, M. & Gillilan, R. E. High-throughput biological small-angle X-ray scattering with a robotically loaded capillary cell. *J. Appl. Crystallogr.* **45,** 213–223 (2012).
29. Petoukhov, M. V. *et al.* New developments in the ATSAS program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* **45,** 342–350 (2012).

Acknowledgements We acknowledge access to the X25 beamline at NSLS and thank the NSLS staff. NSLS is supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. DE-AC02-98CH10886. We thank J. Tainer and J. Perry for their support and access to SIBYLS beamline at the Advanced Light Source, Lawrence Berkeley National Laboratory. We thank E. Arnold, J. Bonanno, S. K. Burley, J. Chiu, D. Comoletti, E. Elrod, F. Jiang, S. Khare, P. Lobel, A. Shatkin, A. Stock, J. Shires and A. Thanou for providing helpful comments and assistance. Special thanks to C. M. Rice for providing J6 HCV clone and support. This work was supported by a Yerkes Research Center Base Grant RR-00165 (A.G.) and NIH grants P50 GM103368 (J.M.), R01 AI080659 (J.M.) AI070101 (A.G.) and DK083356 (A.G.). A.G.K. was supported by a grant from the New Jersey Commission on Cancer Research (DFHS13CRP001).

Author Contributions The project was initiated, designed and supervised by A.G. and J.M. A.G.K., J.W., A.V.Z., A.C. and S.A.Y. designed protein constructs and established purification protocols. A.G.K. prepared all protein crystals. The mouse monoclonal antibody was produced by H.S. and A.G., and was sequenced by C.D.B. and J.J. A.A.P. and A.G. performed the virus neutralization and patient sera ELISA. J.M., M.T.M. and A.G.K. collected, processed and analysed the X-ray crystallographic and SAXS data. A.G.K., J.W. and J.M. wrote the paper with contributions from all authors.

Author Information The coordinates and structure factors have been deposited to the Protein Data Bank under accession code 4NX3. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.M. (jmarco@cabm.rutgers.edu) or A.G. (arash.grakoui@emory.edu).

METHODS

J6 E2 expression. eE2, eE2(ΔHVR1) and E2 core domain encompasses residues 384–656, 413–656 and 456–656 from the HCV J6 genome, respectively. Owing to incomplete deglycosylation at N7 (542) with EndoH, the crystallization construct contained an asparagine to glutamine mutation at this position. The expression constructs consisted of a CMV promoter, a prolactin signal sequence, E2 fragment, PreScission Protease cleavage site and a C-terminal protein-A (ProtA) tag. The entire prolactin-E2-ProtA sequence was PCR amplified and cloned into the pJG lentiviral vector (J. Shires).

Wild-type and GnTI− HEK293T cells[27] (provided by D. Comoletti) were maintained in Dulbecco's Modified Eagle Medium (DMEM) with 10% fetal bovine serum (FBS) at 37 °C with 5% CO_2. One day before the planned transfection, a single T-225 monolayer flask was seeded with 6.0×10^6 HEK293T cells. 90 μg pJG-E2, 60 μg psPAX2 (HIV Gag-Pol packaging vector), 30 μg pMD2.G (VSV glycoprotein vector) and 450 μl of 2 M $CaCl_2$ were mixed and brought to a final volume of 4.5 ml with ddH$_2$0. 4.5 ml of 2× HEPES buffered saline was added at room temperature. After a 2-min incubation, the mixture was added directly to HEK293T cells. After 6–8 h, the media was replaced with DMEM with 10% FBS and 1% antibiotic/antimycotic (A/A) media and incubated for another 70 h.

Two days after transfection, 10,000 GnTI− HEK293T cells were seeded into a single well of a 96-well plate. The supernatant from the transfection, containing the recombinant lentiviruses, was collected and centrifuged for 30 min at 4,000*g* at 4 °C to pellet large cellular debris. Clarified supernatant was transferred to a Beckman Ultracentrifuge tube and virus was pelleted for 1.5 h at 25,000 r.p.m. (80,000*g*) at 4 °C in an SW28 rotor. Supernatant was discarded and the pellet re-suspended in 120 μl of DMEM containing 20% FBS, 1% A/A, and 8 μg ml^{-1} of polybrene. 60 μl of virus suspension was added to the prepared GnTI− HEK293T cells and incubated overnight. Infected cells were expanded and ultimately seeded into an adherent cell bioreactor (Cesco Bioengineering) for long-term growth and protein production.

eE2, eE2(ΔHVR1) and E2 core purification. Cell supernatant containing E2-ProtA was centrifuged for 10 min at 7,000*g*, filtered through a 0.22-μm membrane, and loaded onto an IgG FF column (GE Healthcare). The column was extensively washed with 20 mM sodium phosphate pH 7.0 and then equilibrated with 20 mM HEPES pH 7.5, 250 mM NaCl and 5% glycerol. PreScission Protease was added into the column at approximately 400 μg l^{-1} of supernatant and incubated overnight at 4 °C. For deglycosylation, the pH of the protein solution was adjusted using 1 M sodium citrate pH 5.5 to a final concentration of 100 mM. EndoH was added at a ratio of 1 mg per 2 mg of E2 and the reaction was incubated at room temperature for 3–4 h. The deglycosylated proteins were desalted into 20 mM HEPES pH 7.5, 50 mM NaCl, and 5% glycerol and purified by heparin affinity followed by size-exclusion chromatography over a Superdex200 column. Final yields for all E2 proteins averaged 30 mg l^{-1} of supernatant.

Crystallization. A 1.1:1 molar ratio of E2 core to Fab was incubated for 1–2 h at 4 °C. The complex was purified over a Superdex200 column equilibrated with 20 mM HEPES pH 7.5 and 100 mM NaCl. The complex was concentrated to 5–7 mg ml^{-1} and crystals were grown by the hanging-drop vapour diffusion method. Briefly, 2.5 μl of complex was mixed with an equal volume of reservoir solution, comprising of 22% (w/v) PEG 3350, 0.5 M $MgCl_2$, 0.1 M HEPES pH 7.5, and 15% (v/v) dioxane. Initially, clusters of plate-like crystals grew in 3–4 days. Single, plate-like crystals were obtained via microseeding using a similar reservoir solution supplemented with 2% (v/v) formamide. Crystals were cryoprotected using reservoir solution with 24% (v/v) ethylene glycol and flash cooled in liquid nitrogen. Data were collected at a wavelength of 0.979 Å using beamline X25 of the National Synchrotron Light Source (NSLS), Brookhaven National Laboratory.

Structure determination and refinement. The crystals belong to space group $P2_12_12$ with cell parameters $a = 85.96$ Å, $b = 194.57$ Å, $c = 37.92$ Å. Phases were determined by the molecular replacement method using PHENIX[30] and the coordinates from chains A and B from PDB entry 2GSI. Unambiguous placement of the Fab heavy and light chains provided the necessary phases to extend the map to cover E2 core domain using iterative rounds of model building and density modification by COOT[31], PHENIX, REFMAC[32] and PARROT[33]. The final model was built to a resolution of 2.4 Å, comprising residues 492–522, 538–571 and 596–649 of E2 from the J6 genome, 1–217 of 2A12 light chain, and 1–133 and 136–218 of 2A12 heavy chain with two *N*-linked, *N*-acetylglucosamine, six molecules of formamide and 141 solvents molecules. The model coordinates were refined to R_{work} 0.217 and R_{free} 0.269. Model validation demonstrated 95.0% of the residues located in the most favourable region of the Ramachandran plot with 4.8% in the generously allowed regions[30]. Statistics of the data processing and structure refinement are summarized in Extended Data Table 1.

Small angle X-ray scattering (SAXS). Glycosylated E2 proteins were purified over IgG and anion exchange columns. The proteins were equilibrated with either pH 7.5 buffer (50 mM HEPES pH 7.5, 250 mM NaCl and 1% glycerol) or pH 5.0 buffer (50 mM sodium citrate pH 5.0, 250 mM NaCl and 1% glycerol) by Superdex200 gel filtration column. Glycosylated eE2(ΔHVR1) alone or complex with CD81-LEL (1:2 molar ratios) was purified using pH 7.5 buffer by gel filtration chromatography. Three concentrations of each protein were prepared along with their respective buffers as background control. SAXS data was collected on the SIBYLS beamline at the Advanced Light Source, Lawrence Berkeley National Laboratory. Sample analysis and processing was performed using BioSAXS RAW[28], ATSAS[29], and GNOM[34]. The *ab initio* models were calculated using the application DAMMIF[35]. Consensus models and the normalized spatial discrepancy (NSD) values were calculated by averaging 10 *ab initio* models using the application DAMAVER[36]. X-ray structures of CD81 (PDB ID 1G8Q), HCV E2 core and *ab initio* models were aligned using the application SUPCOMB[37].

Hydrogen deuterium exchange. HD exchange experiments were conducted as described previously[38]. Briefly, 5 μl of deglycosylated eE2 (1.5 mg ml^{-1}), in 200 mM NaCl, 20 mM HEPES pH 7.5, was incubated with 15 μl of the same buffer made with 99.96% ^2H_2O (Cambridge Isotope Laboratories) for 10, 100, or 1,000 s and quenched in 30 μl of 2 M urea, 0.8% formic acid and 50 mM tris(2-carboxyethyl) phosphine (TCEP). The reaction mixture was immediately frozen on dry ice until injection. For the zero time point experiment, the protein was incubated in the buffer made with ^1H_2O and then quenched and frozen. To correct background exchange, a completely deuterated sample was produced by incubating the protein with 100 mM TCEP in 99.96% ^2H_2O overnight before being quenched and frozen. Dionex RSLC with a C18 column (2.1 × 50 mm, 3 μm, Q-C18, 150A, CMP Scientific) and LTQ Velos Orbitrap pro were used for LC–MS analysis. The mass was measured using Orbitrap with resolution of 60,000 and mass range from 300 to 2,000 *m/Z*. The LC–MS data were analysed using HDexaminer 1.2.0 (Serra Analytics) with manual checking of each peptide afterwards.

Limited proteolysis. 8–10 μg of deglycosylated eE2 protein was mixed with trypsin, chymotrypsin or GluC at 1:120 (w/w) ratio (endopeptidase:E2) and incubated at room temperature. Samples were taken at noted time points and analysed by reducing SDS–PAGE, mass spectrometry and N-terminal sequencing.

Production of monoclonal antibody 2A12. Six-to-eight-week-old female BALB/c mice were immunized intraperitoneally with 50 μg eE2 in either complete Freund's adjuvant (first immunization only), or incomplete Freund's adujvant bi-weekly for 8 weeks. A final immunization with 50 μg of eE2 was given intravenously 4 days before collection of splenocytes. Hybridomas were generated using a cloned HAT-sensitive mouse myeloma cell line as a fusion partner. Proliferating hybridomas were screened for their ability to bind eE2 via ELISA, at which point 2A12 was positively identified. Monoclonal antibodies were generated in the laboratory of A. Grakoui (IACUC protocol number YER-2002369-070816GN).

Generation and purification of 2A12 Fab. Hybridoma cells were expanded to a final volume of 2 l in spinner flasks at 100 r.p.m. using Iscove's Modified Dulbecco's Medium, 10% ultra-low IgG FBS, 1% A/A, and 10 mM HEPES (Life Technologies). Cells were collected at $2–3 \times 10^6$ cells per ml, centrifuged for 10 min at 7,000*g*, filtered through a 0.22-μm membrane, and loaded onto a Protein G column (GE Healthcare Life Sciences). After completion, the column was washed with 20 mM sodium phosphate (pH 7.0) followed by phosphate buffered saline (PBS). The antibody was eluted with 0.05% TFA in 2 ml fractions into tubes containing 100 μl of 1 M Tris pH 7.5 for immediate pH neutralization. The eluted antibody was dialysed into 20 mM sodium phosphate pH 7.0 and 10 mM EDTA. Insoluble papain was added to the antibody at 0.15 mg per 1 mg of antibody. Freshly prepared L-cysteine was added to the reaction to a final concentration of 20 mM and mixed at 37 °C for 2 h. The papain was removed by centrifugation at 3,500*g* for 2 min and filtration through a 0.22-μm membrane. Fab was purified by subtractive chromatography over Protein A FF column and desalted into 20 mM Tris pH 8.0.

Sequencing Ig H and L chain gene segments of 2A12 antibody. Total RNA isolated from 2A12 hybridoma cells was reverse transcribed into cDNA using random hexamers. Expressed heavy (H) and light (L) chains were amplified using standard primers that are complimentary to all murine H and L chain gene segments[39]. The PCR products were sequenced either directly or following cloning into pCR 2.1-TOPO vector (Life Technologies).

CD81 purification and binding assays. Human CD81-LEL (residues 112–202) was produced as a fusion with C-terminal ProtA tag in HEK293T cells using the same lentiviral expression system described for eE2. Cell culture supernatants were loaded onto an IgG FF column, washed with 20 mM sodium phosphate pH 7.0, eluted with 100 mM sodium citrate pH 3.0 containing 20 mM KCl and immediately neutralized with 1 M Tris pH 9.0. The ProtA tag was cleaved by PreScission Protease in a ratio of 1:50 (w/w) followed by overnight dialysis in 20 mM HEPES pH 7.5, 250 mM NaCl, and 5% glycerol. High-purity CD81 protein was obtained by anion exchange and size-exclusion chromatography.

For binding studies, a 96-well plate (Nalgene Nunc, Thermo Fisher Scientific) was coated with 50 μg of CD81-LEL overnight at 4 °C. All experiments were

duplicated against BSA as a negative control. Plates were washed three times with PBS containing 0.05% Tween 20 (PBS-T) and blocked with 3% (w/v) BSA in PBS-T for 1 h at room temperature. 50 μl of eE2 or E2 core at different concentrations was added to appropriate wells and incubated overnight at 4 °C. On day 3, the wells were washed three times with PBS-T and incubated with monoclonal antibody 2A12 cell supernatant for 1 h at room temperature. Plates were washed three times with PBS-T and incubated with anti-mouse-HRP conjugated antibody for 1 h at room temperature. Finally, the plate was washed five times with PBS-T. 50 μl of TMB substrate (ThermoFisher Scientific) was added to each well and incubated for 5 min, followed by the addition of 50 μl of 2 M sulphuric acid to stop the reaction. Absorbance readings were acquired at 450 nm using Softmax Pro software on a Spectra Max 250 (Molecular Devices).

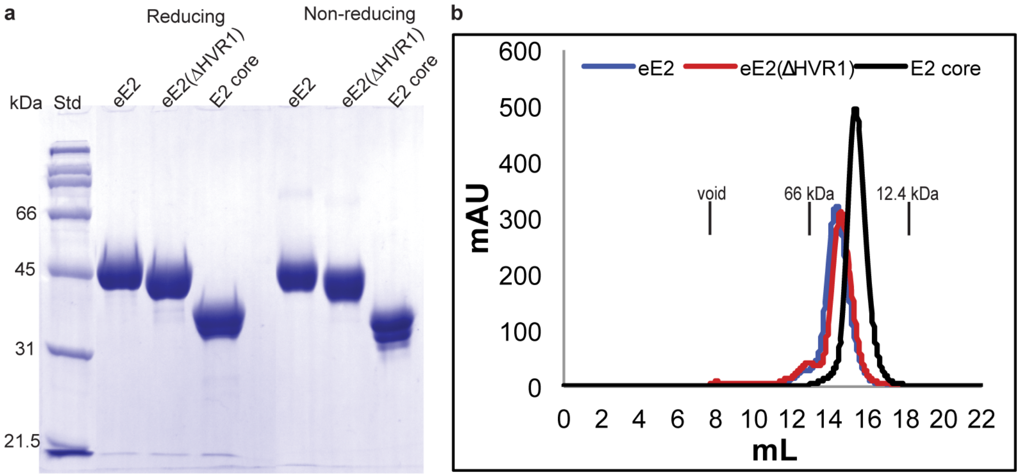
Neutralization assay. Huh-7.5 cells were maintained in DMEM containing 10% FBS (Hyclone) and 100 $\mu g\,ml^{-1}$ of penicillin and streptomycin (Cellgro) at 37 °C in 5% CO_2. Naive Huh-7.5 cells were seeded at 6,000 cells per well in a 96-well plate. The following day, 100 μl of 2C1, 2A12, or H113 serially diluted in complete media were added per well at various concentrations. In parallel, 100 μl of eE2, E2 core, gp140, or CD81-LEL serially diluted in complete DMEM were added at varying concentrations beginning at 100 $\mu g\,ml^{-1}$. Cells were then infected with 100 μl of genotype 2a virus Cp7 encoding the *Renilla* luciferase gene[40]. 72 h after infection, relative light units were measured on a Clarity 4.0 luminometer (Biotek) using the *Renilla* Luciferase Assay System (Promega).

Assessment of cellular cytotoxicity. Huh-7.5 cells were incubated with varying concentrations of protein as described above, beginning at 100 $\mu g\,ml^{-1}$. After 72 h, cells were washed once with PBS, treated with trypsin, and collected in 100 μl PBS. Cells were stained with 7-AAD according to the manufacturer's instructions (BD Biosciences) and analysed using a BD LSR II and FlowJo software (Tree Star).

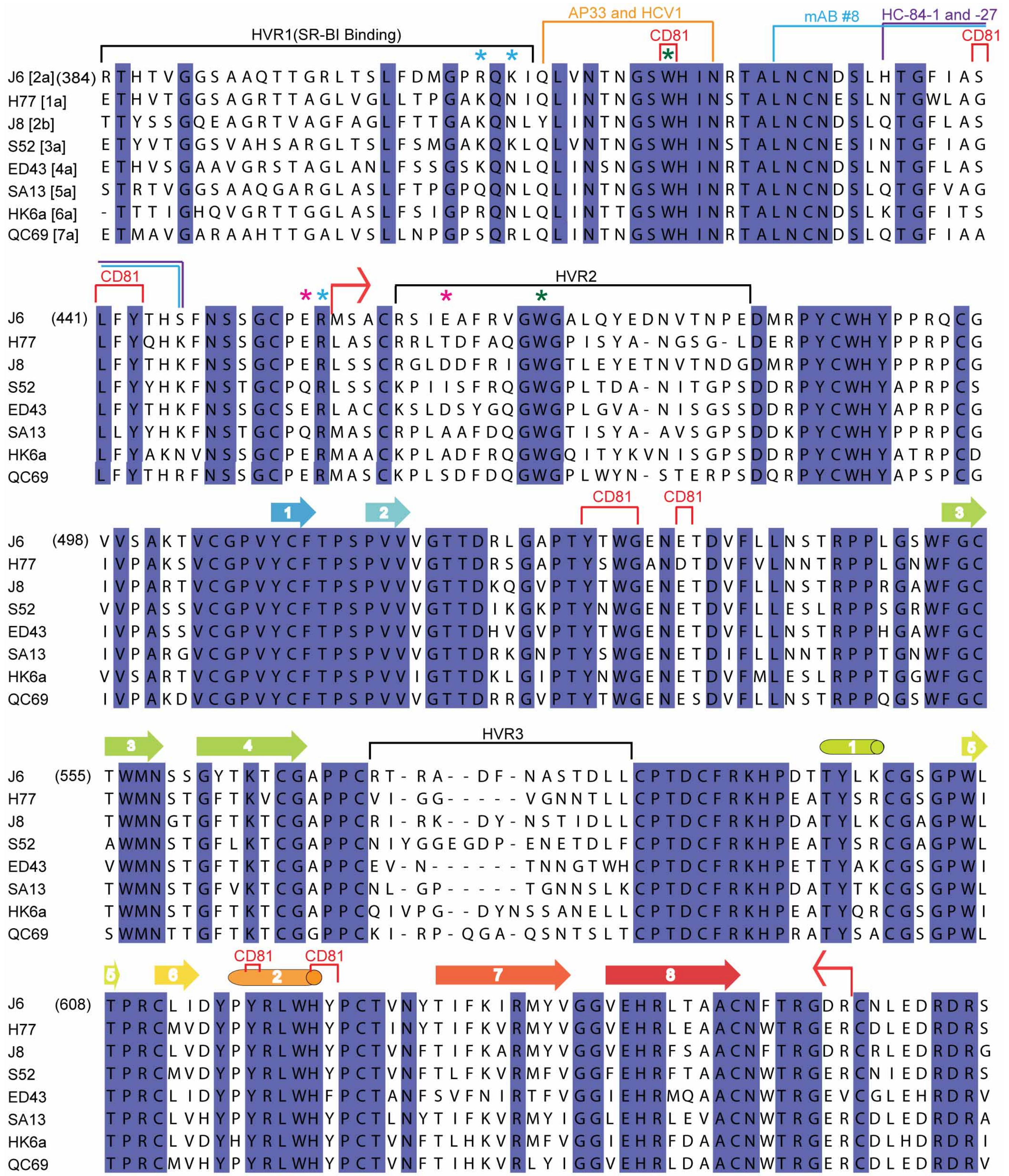
Human plasma ELISA. 96-well enzyme immunoassay plates (ThermoFisher Scientific) were coated overnight at 4 °C with 50 μl of a 1 $\mu g\,ml^{-1}$ solution of eE2 or E2 core diluted in 0.1 M Na_2CO_3. Plates were washed twice with PBS-T and then blocked for 1 h at 37 °C in PBS-T containing 10% fetal calf serum (HyClone). Blood samples were collected in heparin tubes (Becton Dickinson) and plasma was isolated and frozen at −80 °C. Plasma was serially diluted in a binding solution composed of 0.1% (v/v) normal goat serum in PBS-T (Jackson ImmunoResearch Laboratories). 100 μl of sample were added per well and incubated at room temperature for 90 min. After eight washes with PBS, 100 μl of mouse anti-human IgG biotin antibody (Mabtech) diluted 1:20,000 in binding solution were added per well and incubated 1 h at room temperature. Following five additional washes with PBS, 100 μl streptavidin-horseradish peroxidase (HRP) conjugate was added to each well at a 1:2,000 dilution in binding buffer and incubated for 45 min at room temperature (Mabtech). Absorbance was measured and analysed using a VersaMax microplate reader and SoftMax Pro software (Molecular Devices) following five washes and the addition of tetramethylbenzidine substrate solution (Ebioscience). Human sera were isolated from whole-blood samples, and informed consent was obtained for all subjects (IRB no. 1358-2004, Emory University School of Medicine, principal investigator A. Grakoui).

Alignment. Secondary structures were assigned using the program DSSP[41]. Sequences were obtained from the National Center for Biotechnology Information (NCBI) using the following accession numbers: J6 ADV40003.1, H77 ACA53555.1, J8 P26661.3, S52 AEB71616.2, ED43 AEB71617.2, SA13 AEB71618.2, HK6a AEB71619.2, QC69 ACM69041.1. The E2 sequences were aligned with multiple alignment using fast Fourier transform (MAFFT)[42] and edited for figure generation using JalView version 2 (ref. 43).

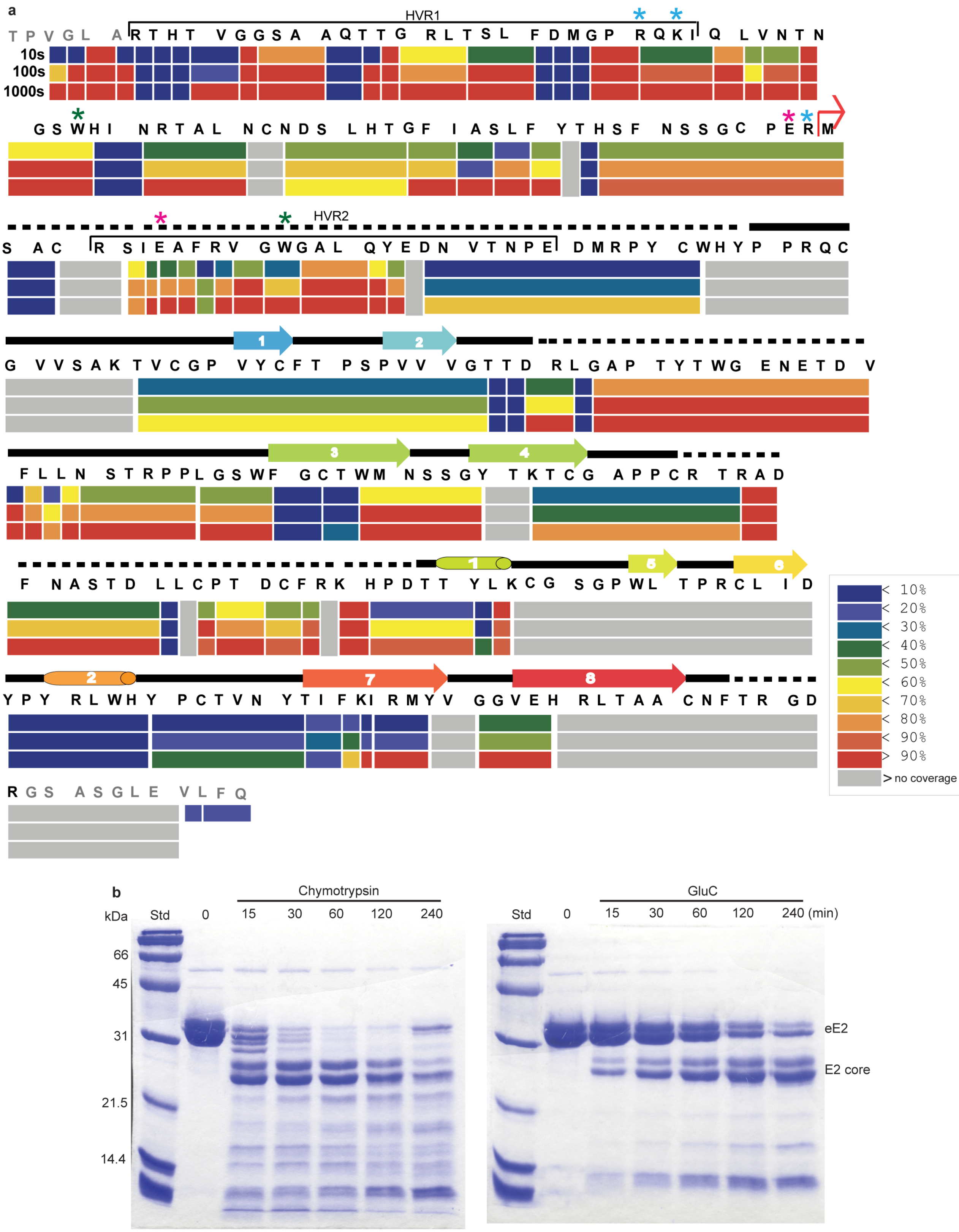
30. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66,** 213–221 (2010).
31. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66,** 486–501 (2010).
32. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D* **67,** 355–367 (2011).
33. Zhang, K. Y., Cowtan, K. & Main, P. Combining constraints for electron-density modification. *Methods Enzymol.* **277,** 53–64 (1997).
34. Semenyuk, A. V. & Svergun, D. I. GNOM - a program package for small-angle scattering data processing. *J. Appl. Crystallogr.* **24,** 537–540 (1991).
35. Svergun, D. I. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **76,** 2879–2886 (1999).
36. Volkov, V. V. & Svergun, D. I. Uniqueness of *ab initio* shape determination in small-angle scattering. *J. Appl. Crystallogr.* **36,** 860–864 (2003).
37. Kozin, M. B. & Svergun, D. Automated matching of high- and low-resolution structural models. *J. Appl. Crystallogr.* **34,** 33–41 (2001).
38. Sharma, S. *et al.* Construct optimization for protein NMR structure analysis using amide hydrogen/deuterium exchange mass spectrometry. *Proteins* **76,** 882–894 (2009).
39. Tiller, T., Busse, C. E. & Wardemann, H. Cloning and expression of murine Ig genes from single B cells. *J. Immunol. Methods* **350,** 183–193 (2009).
40. Mateu, G., Donis, R. O., Wakita, T., Bukh, J. & Grakoui, A. Intragenotypic JFH1 based recombinant hepatitis C virus produces high levels of infectious particles but causes increased cell death. *Virology* **376,** 397–407 (2008).
41. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22,** 2577–2637 (1983).
42. Katoh, K., Kuma, K., Toh, H. & Miyata, T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* **33,** 511–518 (2005).
43. Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M. & Barton, G. J. Jalview Version 2–a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25,** 1189–1191 (2009).



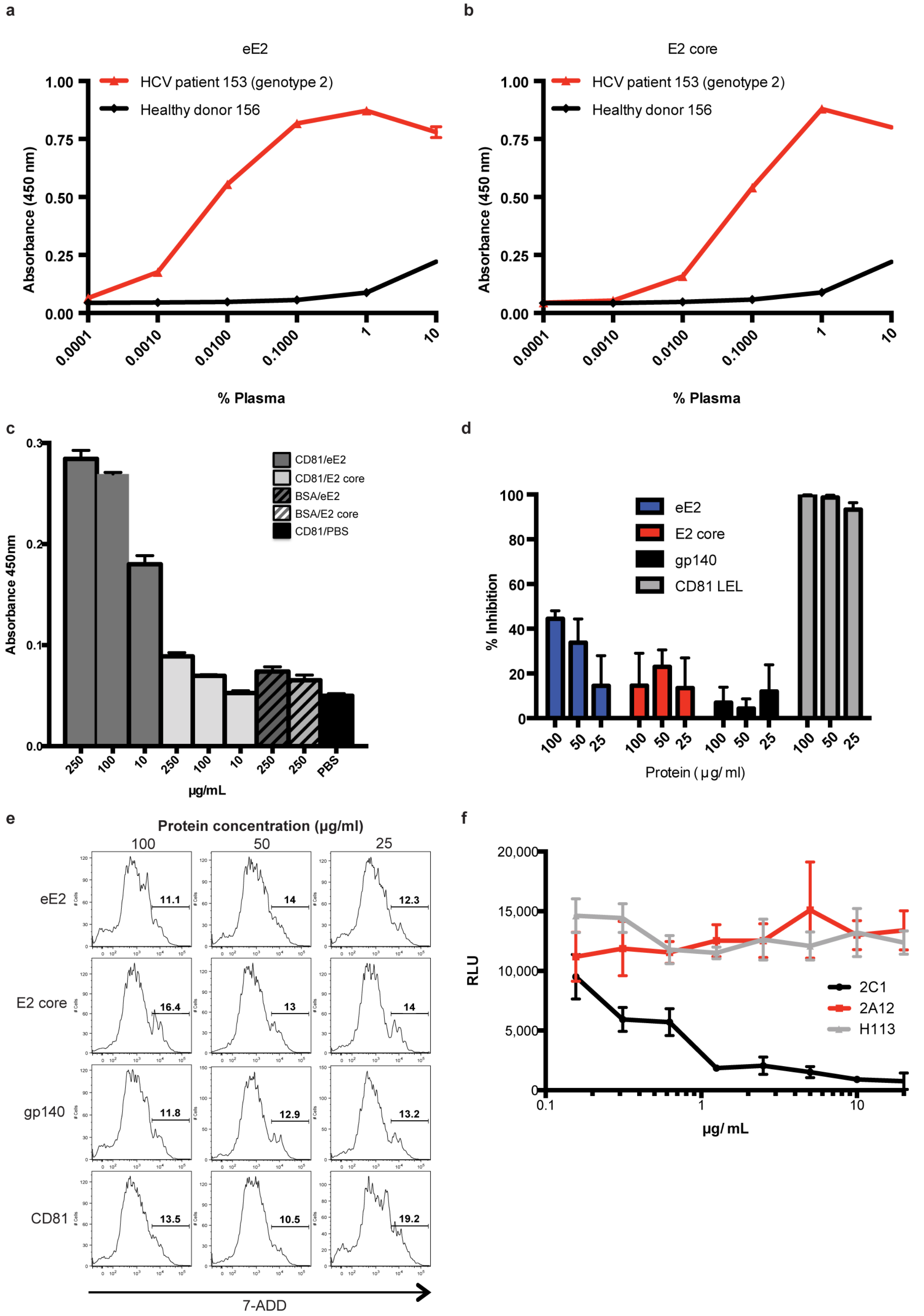
Extended Data Figure 1 | eE2, eE2(ΔHVR1) and E2 core are highly soluble and monomeric in solution. **a**, A comparison of proteins under reducing and non-reducing conditions is shown by a 10% SDS–PAGE gel with protein standards (Std). **b**, Size-exclusion chromatography of eE2, eE2(ΔHVR1) and E2 core proteins on a Superdex200 gel filtration column. The elution positions of the void volume (>200 kDa), albumin (66 kDa) and cytochrome C (12.4 kDa) are indicated. Molecular masses of eE2, eE2(ΔHVR1) and E2 core are ~46 kDa, ~42 kDa and ~32 kDa, respectively.



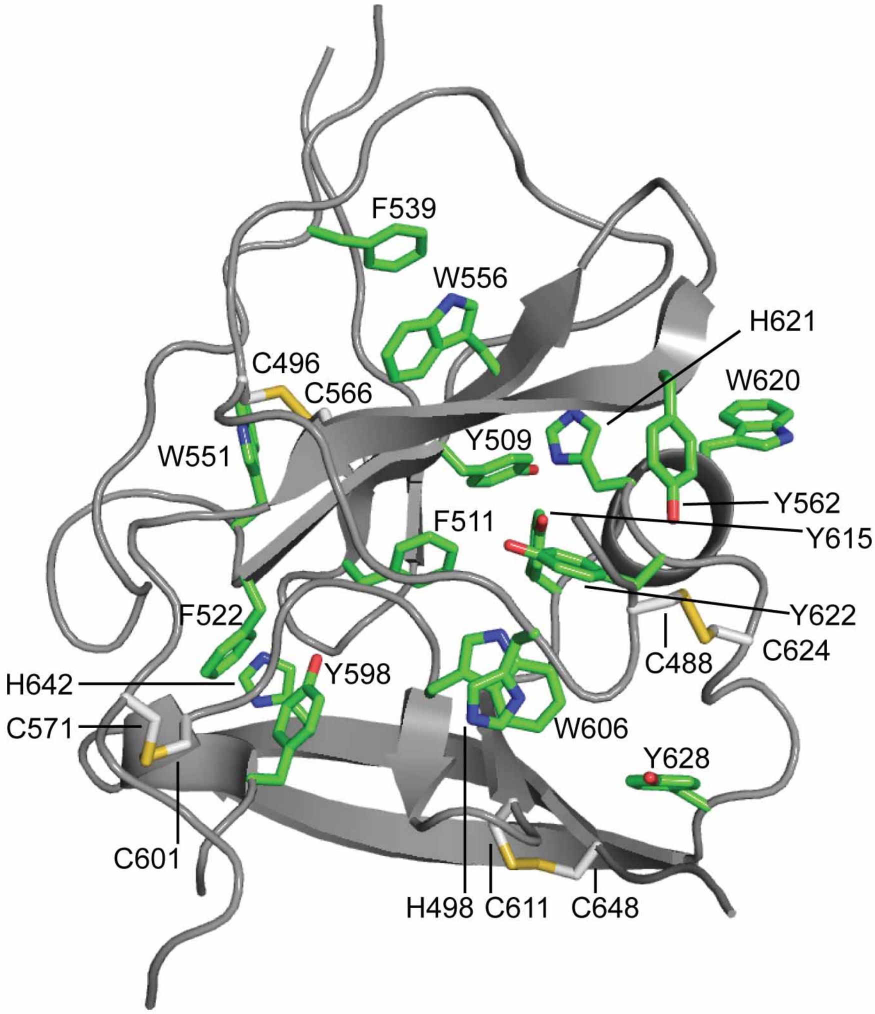
Extended Data Figure 2 | eE2 sequence alignment. Red bent arrows indicate the N- and C-terminal boundaries of the E2 crystallization construct. Cylinders and arrows represent α-helices and β-strands, respectively, and are coloured according to cartoon representation in Fig. 1. CD81 binding regions are bracketed in red; hypervariable regions are bracketed in black. SR-BI binds to HVR1. The asterisks indicate the location of trypsin (blue), chymotrypsin (green) and GluC (magenta) cleavage sites. The binding sites of neutralizing antibodies for which structural information is available are coloured orange for HCV1 and AP33, blue for mAb 8, and purple for HC34-1 and HC34-17.



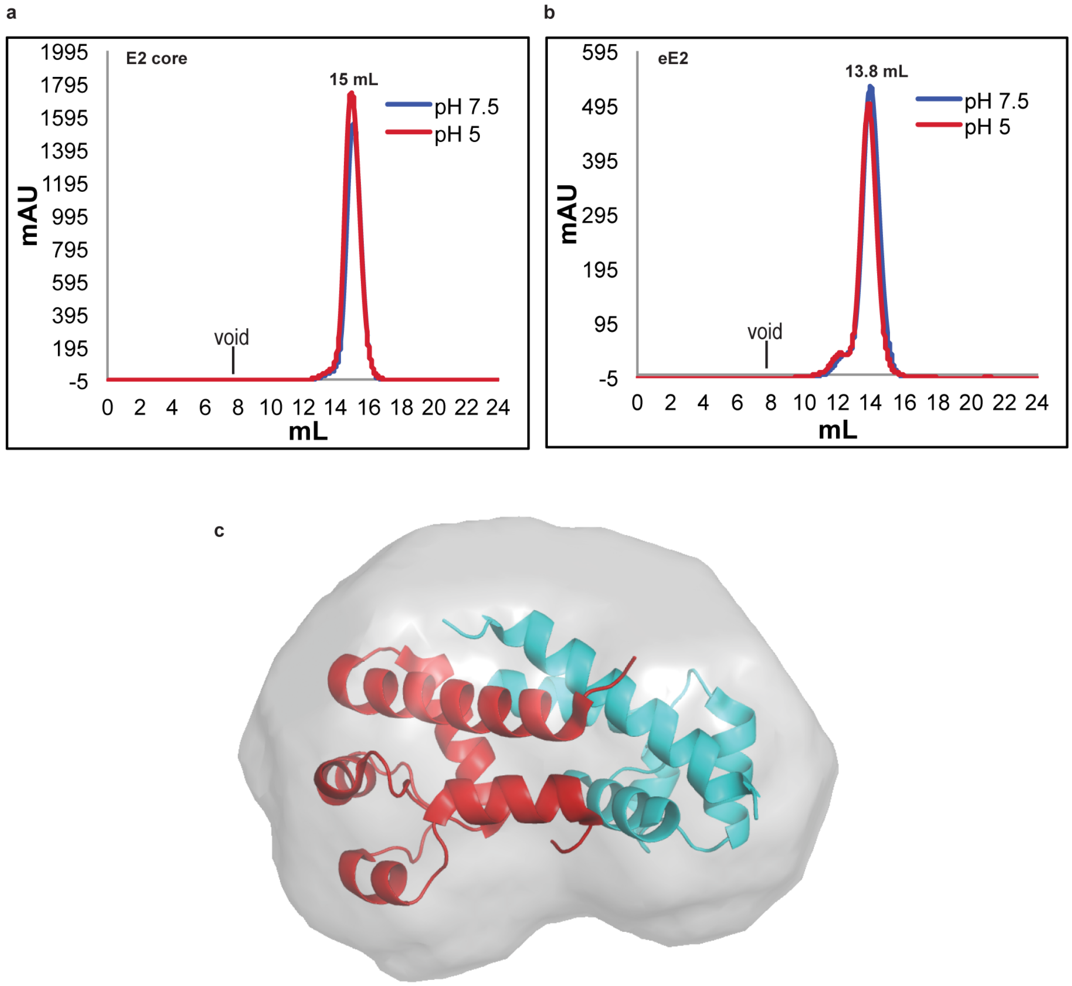
Extended Data Figure 3 | Hydrogen deuterium exchange and limited proteolysis of eE2. **a**, The percentage hydrogen deuterium exchange shown at 10, 100 and 1,000 s time points. The secondary structure of E2 core is placed above to emphasize flexible regions. A red arrow indicates the E2 core N terminus. Extra residues (grey) on N and C terminus come from the vector. Potential cleavage sites for trypsin (blue), chymotrypsin (green) and GluC (magenta) are indicated by asterisks. The colour pattern indicates the percentage of exchange. Grey areas are the regions of no coverage. **b**, Digestion of deglycosylated eE2 with chymotrypsin (left) and GluC (right) reveals a shift from the ~35-kDa untreated protein (0 min) to ~25 kDa after digestion. Samples were taken at the indicated time points and analysed by reducing 12% SDS–PAGE gel. Molecular mass protein standards (Std) are indicated. The bands were analysed by N-terminal sequencing and mass spectrometry.



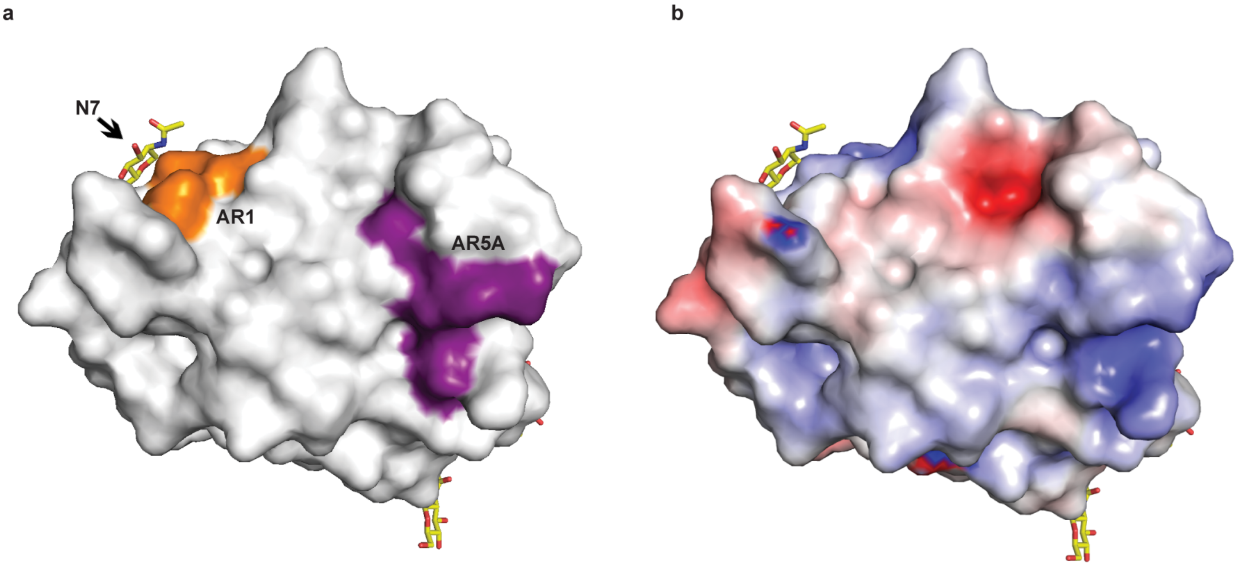
Extended Data Figure 4 | Functional analyses of eE2 and E2 core. **a**, Antibodies from patient sera infected with HCV genotype 2 show a concentration-dependent binding to eE2 (red) whereas healthy donor sera exhibit only background binding (black). **b**, Similar binding is observed for E2 core. The measurements were done in triplicate with the error bars representing the standard error of the mean (s.e.m.). **c**, E2 core (light grey) shows reduced binding to CD81 when compared to eE2 (dark grey) by an ELISA. Bars with stripes indicate E2 binding to a negative control, BSA. The solid black bar indicates CD81 binding to PBS, used to verify the absence of background. The measurements were done in triplicate with the error bars representing the s.e.m. **d**, eE2 (blue) and CD81-LEL (positive control, grey) inhibit the infection. E2 core (red) shows reduced inhibition. HIV gp140 (black) expressed in the same system was used as a negative control. The measurements were done in triplicate with the error bars representing the s.e.m. **e**, To rule out the possibility of toxic effects from the recombinant proteins, the cell viability was measured as described in Methods, using similar protein concentrations as in **d**. **f**, In an ELISA, 2A12 (red), and an irrelevant antibody, H113 (grey), fail to neutralize HCVcc infection. 2C1 (positive control, black), a mouse monoclonal antibody that binds to the disordered N-terminal region of eE2, blocks infection. The measurements were done in triplicate with the error bars representing the s.e.m.



Extended Data Figure 5 | E2 core contains an extensive hydrophobic core. Sheets A and B are held together by an extensive hydrophobic core composed of mostly aromatic amino acids (green) and five disulphide bonds (yellow).



Extended Data Figure 6 | eE2 and E2 core do not undergo oligomeric changes at low pH. a, b, An overlay of E2 core (**a**) and eE2 (**b**) elution profiles from Superdex200 gel filtration at pH 7.5 (blue) and pH 5.0 (red). The expected void volume and observed elution positions of individual proteins are indicated. **c**, The SAXS envelope of CD81-LEL fit with a dimer crystal structure (PDB 1G8Q). The individual proteins of the CD81-LEL dimer are coloured red and blue.



Extended Data Figure 7 | Epitope mapping of conformational antibodies on E2 core surface. **a**, Surface epitopes of AR1 (orange) are shown. AR1 blocks the E1E2 heterodimer binding to CD81. AR5A (purple) inhibits E1E2 heterodimerization and is mapped on a well-conserved hydrophobic surface of the core. **b**, Surface of E2 core coloured by electrostatic potential. The view in **a** and **b** is identical.

Extended Data Table 1 | Summary of the X-ray crystallographic analyses

Data collection	
Wavelength	0.979 Å
Space group	$P2_12_12$
Cell dimensions	
a, *b*, *c* (Å)	85.96, 194.57, 37.92
α, *β*, *γ* (°)	90, 90, 90
Resolution (Å)	24.69-2.40 (2.49-2.40)*
R_{merge}	0.118 (1.042)
R_{pim}	0.054 (0.490)
I/σ	9.6 (1.9)
Completeness (%)	100 (100)
Redundancy	5.7 (5.3)
Refinement	
Resolution (Å)	24.69-2.40
No. unique reflections	24349 (2517)
R_{work}/R_{free}	0.217/0.269
No. atoms	
Protein/glycans	4162
Solvent	137
Ions	18
B-factors (Å^2)	
Protein/glycan	54
Solvent	45
Ions	47
R.M.S deviations	
Bond lengths (Å)	0.006
Bond angles (º)	1.14
Ramachandran favored (%)	95.0
Ramachandran outliers (%)	0.2

* Highest resolution shell is shown in parenthesis.

Extended Data Table 2 | Summary of SAXS analyses

Protein	R_g (Å)	D_{max} (Å)	NSD
E2 core	27.8±0.03	95	0.705±0.026
eE2(ΔHVR1)	28.9±0.05	101	0.696±0.075
eE2	28.2±0.02	84	0.702±0.054
eE2(ΔHVR1) + CD81	36.8±0.22	127	0.714±0.026
CD81	18.8±0.04	64	0.554±0.007
E2 core pH 5.0	27.8±0.07	95	-
eE2 pH 5.0	27.8±0.07	95	-

doi:10.1038/nature13314

A semi-synthetic organism with an expanded genetic alphabet

Denis A. Malyshev[1], Kirandeep Dhami[1], Thomas Lavergne[1], Tingjian Chen[1], Nan Dai[2], Jeremy M. Foster[2], Ivan R. Corrêa Jr[2] & Floyd E. Romesberg[1]

Organisms are defined by the information encoded in their genomes, and since the origin of life this information has been encoded using a two-base-pair genetic alphabet (A–T and G–C). *In vitro*, the alphabet has been expanded to include several unnatural base pairs (UBPs)[1–3]. We have developed a class of UBPs formed between nucleotides bearing hydrophobic nucleobases, exemplified by the pair formed between d5SICS and dNaM (d5SICS–dNaM), which is efficiently PCR-amplified[1] and transcribed[4,5] *in vitro*, and whose unique mechanism of replication has been characterized[6,7]. However, expansion of an organism's genetic alphabet presents new and unprecedented challenges: the unnatural nucleoside triphosphates must be available inside the cell; endogenous polymerases must be able to use the unnatural triphosphates to faithfully replicate DNA containing the UBP within the complex cellular milieu; and finally, the UBP must be stable in the presence of pathways that maintain the integrity of DNA. Here we show that an exogenously expressed algal nucleotide triphosphate transporter efficiently imports the triphosphates of both d5SICS and dNaM (d5SICSTP and dNaMTP) into *Escherichia coli*, and that the endogenous replication machinery uses them to accurately replicate a plasmid containing d5SICS–dNaM. Neither the presence of the unnatural triphosphates nor the replication of the UBP introduces a notable growth burden. Lastly, we find that the UBP is not efficiently excised by DNA repair pathways. Thus, the resulting bacterium is the first organism to propagate stably an expanded genetic alphabet.

To make the unnatural triphosphates available inside the cell, we previously suggested using passive diffusion of the free nucleosides into the cytoplasm followed by their conversion to the corresponding triphosphate via the nucleoside salvage pathway[8]. Although we have shown that analogues of d5SICS and dNaM are phosphorylated by the nucleoside kinase from *Drosophila melanogaster*[8], monophosphate kinases are more specific[9], and in *E. coli* we found that overexpression of the endogenous nucleoside diphosphate kinase results in poor growth. As an alternative, we focused on the nucleotide triphosphate transporters (NTTs) of obligate intracellular bacteria and algal plastids[10–14]. We expressed eight different NTTs in *E. coli* C41(DE3)[15–17] and measured the uptake of [α-^{32}P]-dATP as a surrogate for the unnatural triphosphates (Extended Data Fig. 1). We confirmed that [α-^{32}P]-dATP is efficiently transported into cells by the NTTs from *Phaeodactylum tricornutum* (*Pt*NTT2)[18] and *Thalassiosira pseudonana* (*Tp*NTT2)[18]. Although NTTs from *Protochlamydia amoebophila* (*Pam*NTT2 and *Pam*NTT5)[15] also import [α-^{32}P]-dATP, *Pt*NTT2 showed the most activity, and both it and *Tp*NTT2 are known to have broad specificity[18], making them the most promising NTTs for further characterization.

Transport via an NTT requires that the unnatural triphosphates are sufficiently stable in culture media; however, preliminary characterization of d5SICSTP and dNaMTP indicated that decomposition occurs in the presence of actively growing *E. coli* (Extended Data Fig. 2). Similar behaviour was observed with [α-^{32}P]-dATP, and the dephosphorylation products detected by thin-layer chromatography (TLC) for [α-^{32}P]-dATP, or by high-performance liquid chromatography (HPLC) and matrix-assisted laser desorption/ionization (MALDI) for d5SICSTP and dNaMTP, suggest that decomposition is mediated by phosphatases. As no degradation was observed upon incubation in spent media, decomposition seems to occur within the periplasm. No increase in stability was observed in cultures of single-gene-deletion mutants of *E. coli* BW25113 lacking a specific periplasmic phosphatase[19] (as identified by the presence of a Sec-type amino-terminal leader sequence), including *phoA*, *ushA*, *appA*, *aphA*, *yjjX*, *surE*, *yfbR*, *yjjG*, *yfaO*, *mutT*, *nagD*, *yggV*, *yrfG* or *ymfB*, suggesting that decomposition results from the activity of multiple phosphatases. However, the extracellular stability of [α-^{32}P]-dATP was significantly greater when 50 mM potassium phosphate (KPi) was added to the growth medium (Extended Data Fig. 3). Thus, we measured [α-^{32}P]-dATP uptake from media containing 50 mM KPi after induction of the transporter with isopropyl-β-D-thiogalactoside (IPTG) (Extended Data Fig. 4). Although induction with 1 mM IPTG resulted in slower growth, consistent with the previously reported toxicity of NTTs[17], it also resulted in maximal [α-^{32}P]-dATP uptake. Thus, after addition of 1 mM IPTG, we analysed the extracellular and intracellular stability of [α-^{32}P]-dATP as a function of time (Extended Data Fig. 5). Cells expressing *Pt*NTT2 were found to have the highest levels of intracellular [α-^{32}P]-dATP, and although both extra- and intracellular dephosphorylation was still observed, the ratio of triphosphate to dephosphorylation products inside the cell remained roughly constant, indicating that the extracellular concentrations and *Pt*NTT2-mediated influx are sufficient to compensate for intracellular decomposition.

Likewise, we found that the addition of KPi increased the extracellular stability of d5SICSTP and dNaMTP (Extended Data Fig. 2), and

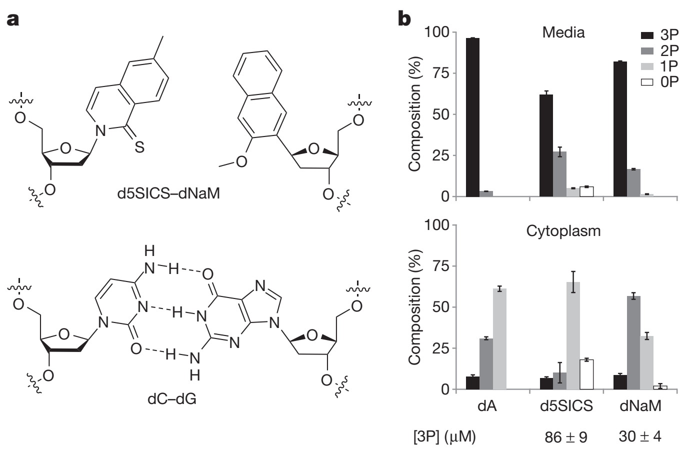


Figure 1 | Nucleoside triphosphate stability and import. **a**, Chemical structure of the d5SICS–dNaM UBP compared to the natural dG–dC base pair. **b**, Composition analysis of d5SICS and dNaM in the media (top) and cytoplasmic (bottom) fractions of cells expressing *Pt*NTT2 after 30 min incubation; dA shown for comparison. 3P, 2P, 1P and 0P correspond to triphosphate, diphosphate, monophosphate and nucleoside, respectively; [3P] is the intracellular concentration of triphosphate. Error bars represent s.d. of the mean, $n = 3$.

[1]Department of Chemistry, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA. [2]New England Biolabs, 240 County Road, Ipswich, Massachusetts 01938, USA.

when a stationary phase culture was diluted 100-fold into fresh media, the half-lives of both unnatural triphosphates (initial concentrations of 0.25 mM) were found to be approximately 9 h, which seemed sufficient for our purposes. Thirty minutes after their addition to the media, neither of the unnatural triphosphates was detected in cells expressing *Tp*NTT2; in contrast, 90 μM of d5SICSTP and 30 μM of dNaMTP were found in the cytoplasm of cells expressing *Pt*NTT2 (Fig. 1b). Although intracellular decomposition was still apparent, the intracellular concentrations of intact triphosphate are significantly above the sub-micromolar K_M values of the unnatural triphosphates for DNA polymerases[20], setting the stage for replication of the UBP in a living bacterial cell.

The replication of DNA containing d5SICS-dNaM has been validated *in vitro* with different polymerases, primarily family A polymerases, such as the Klenow fragment of *E. coli* DNA polymerase I (pol I)[20,21]. As the majority of the *E. coli* genome is replicated by pol III, we engineered a plasmid to focus replication of the UBP to pol I. Plasmid pINF (the information plasmid) was constructed from pUC19 using solid-phase DNA synthesis and circular-extension PCR to replace the dA–dT pair at position 505 with dNaM paired opposite an analogue of d5SICS (dTPT3[22]) (Fig. 2a, b). This positions the UBP 362 bp downstream of the ColE1 origin of replication where leading-strand replication is mediated by pol I[23], and within the TK-1 Okazaki processing site[24], where lagging-strand synthesis is also expected to be mediated by pol I. Synthetic pINF was constructed using the d5SICS analogue because it should be efficiently replaced by d5SICS if replication occurs *in vivo*, making it possible to differentiate *in vivo* replicated pINF from synthetic pINF.

To determine whether *E. coli* can use the imported unnatural triphosphates to stably propagate pINF, C41(DE3) cells were first transformed with a pCDF-1b plasmid encoding *Pt*NTT2 (hereafter referred to as pACS, for accessory plasmid, Fig. 2a) and grown in media containing 0.25 mM of both unnatural triphosphates, 50 mM KPi and 1 mM IPTG to induce transporter production. Cells were then transformed with pINF, and after a 1-h recovery period, cultures were diluted tenfold with the same media supplemented with ampicillin, and growth was monitored via culture turbidity (Extended Data Table 1). As controls, cells were also transformed with pUC19, or grown without either IPTG or without the unnatural triphosphates. Again, growth was significantly slower in the presence of IPTG, but the addition of d5SICSTP and dNaMTP resulted in only a slight further decrease in growth in the absence of pINF, and interestingly, it eliminated a growth lag in the presence of pINF (Fig. 2c), suggesting that the unnatural triphosphates are not toxic and are required for the efficient replication of pINF.

To demonstrate the replication of pINF, we recovered the plasmid from cells after 15 h of growth. The introduction of the UBP resulted in

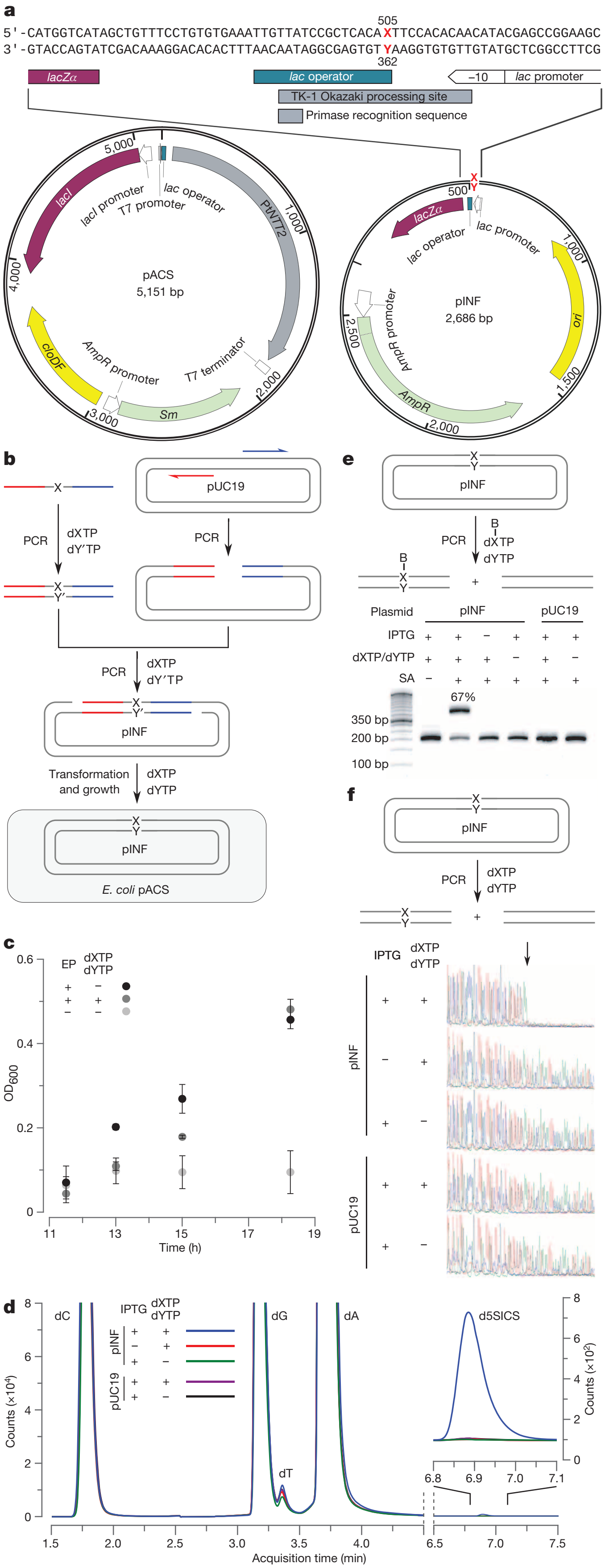


Figure 2 | Intracellular UBP replication. a, Structure of pACS and pINF. dX and dY correspond to dNaM and a d5SICS analogue[22] that facilitated plasmid construction (see Methods). *cloDF*, origin of replication; *Sm*, streptomycin resistance gene; *AmpR*, ampicillin resistance gene; *ori*, ColE1 origin of replication; *lacZα*, β-galactosidase fragment gene. **b**, Overview of pINF construction. A DNA fragment containing the unnatural nucleotide was synthesized via solid-phase DNA synthesis and then used to assemble synthetic pINF via circular-extension PCR[29]. X, dNaM; Y′, dTPT3 (an analogue of d5SICS[22]); y, d5SICS (see text). Colour indicates regions of homology. The doubly nicked product was used directly to transform *E. coli* harbouring pACS. **c**, The addition of d5SICSTP and dNaMTP eliminates a growth lag of cells harbouring pINF. EP, electroporation. Error bars represent s.d. of the mean, $n = 3$. **d**, LC-MS/MS total ion chromatogram of global nucleoside content in pINF and pUC19 recorded in dynamic multiple reaction monitoring (DMRM) mode. pINF and pUC19 (control) were propagated in *E. coli* in the presence or absence of unnatural triphosphates, and with or without *Pt*NTT2 induction. The inset shows a 100-fold expansion of the mass-count axis in the d5SICS region. **e**, Biotinylation only occurs in the presence of the UBP, the unnatural triphosphates and transporter induction. After growth, pINF was recovered, and a 194-nucleotide region containing the site of UBP incorporation (nucleotides 437–630) was amplified and biotinylated. B, biotin; SA, streptavidin. The natural pUC19 control plasmid was prepared identically to pINF. A 50-bp DNA ladder is shown to the left. **f**, Sequencing analysis demonstrates retention of the UBP. An abrupt termination in the Sanger sequencing reaction indicates the presence of UBP incorporation (site indicated with arrow).

a small (approximately twofold) reduction in the copy number of pINF, as gauged by its ratio to pACS (Extended Data Table 1); we determined that the plasmid was amplified 2×10^7-fold during growth (approximately 24 doublings) based on the amount of recovered plasmid and the transformation efficiency. To determine the level of UBP retention, the recovered plasmid was digested, dephosphorylated to single nucleosides, and analysed by liquid chromatography-tandem mass spectrometry (LC-MS/MS)[25]. Although the detection and quantification of dNaM were precluded by its poor fragmentation efficiency and low product ion counts over background, signal for d5SICS was clearly observable (Fig. 2d). External calibration curves were constructed using the unnatural nucleoside and validated by determining its ratio to dA in synthetic oligonucleotides (Extended Data Table 2). Using the resulting calibration curve, we determined the ratio of dA to d5SICS in recovered pINF was 1,106 to 1, which when compared to the expected ratio of 1,325 to 1, suggests the presence of approximately one UBP per plasmid. No d5SICS was detected in control experiments in which the transporter was not induced, or when the unnatural triphosphates were not added to the media, or when pUC19 was used instead of pINF (Fig. 2d, inset), demonstrating that its presence results from the replication of the UBP and not from misinsertion of the unnatural triphosphates opposite a natural nucleotide. Importantly, as the synthetic pINF contained an analogue of d5SICS, and d5SICS was only provided as a triphosphate added to the media, its presence in pINF confirms *in vivo* replication.

To independently confirm and quantify the retention of the UBP in the recovered plasmid, the relevant region was amplified by PCR in the presence of d5SICSTP and a biotinylated dNaMTP analogue[4] (Fig. 2e). Analysis by streptavidin gel shift showed that 67% of the amplified DNA contained biotin. No shift was observed in control experiments where the transporter was not induced, or when unnatural triphosphates were not added, or when pUC19 was used instead of pINF, demonstrating that the shift results from the presence of the UBP. Based on a calibration curve constructed from the shifts observed with the amplification products of controlled mixtures of DNA containing dNaM or its fully natural counterpart (Methods and Extended Data Fig. 6), the observed gel shift corresponds to a UBP retention of 86%. Similarly, when the amplification product obtained with d5SICSTP and dNaMTP was analysed by Sanger sequencing in the absence of the unnatural triphoshates[1,26,27], the sequencing chromatogram showed complete termination at the position of UBP incorporation, which with an estimated lower limit of read-through detection of 5%, suggests a level of UBP retention in excess of 95% (Fig. 2f). In contrast, amplification products obtained from pINF recovered from cultures grown without *Pt*NTT2 induction, without added unnatural triphosphates, or obtained from pUC19 propagated under identical conditions, showed no termination. Overall, the data unambiguously demonstrate that DNA containing the UBP was replicated *in vivo* and allow us to estimate that replication occurred with fidelity (retention per doubling) of at least 99.4% (24 doublings; 86% retention; $0.994^{24} = 0.86$). This fidelity corresponds to an error rate of approximately 10^{-3}, which is comparable to the intrinsic error rate of some polymerases with natural DNA[28].

The high retention of the UBP over a 15-h period of growth (approximately 24 doublings) strongly suggests that it is not efficiently excised by DNA repair pathways. To test further this hypothesis and to examine retention during prolonged stationary phase growth, we repeated the experiments, but monitored UBP retention, cell growth and unnatural triphosphate decomposition for up to 6 days without providing any additional unnatural triphosphates (Fig. 3 and Extended Data Fig. 7). At 15 and 19 h of growth, the cultures reached an optical density at 600 nm (OD_{600}) of approximately 0.9 and 1.2, respectively, and both d5SICSTP and dNaMTP decomposed to 17–20% and 10–16% of their initial 0.25-mM concentrations (Extended Data Fig. 7a). In agreement with the experiments described above, retention of the UBP after 15 h was $97 \pm 5\%$ and $>95\%$, as determined by gel shift and sequencing, respectively, and after 19 h it was $91 \pm 3\%$ and $>95\%$. As the cultures entered stationary phase and the triphosphates decomposed completely, plasmid loss began to compete

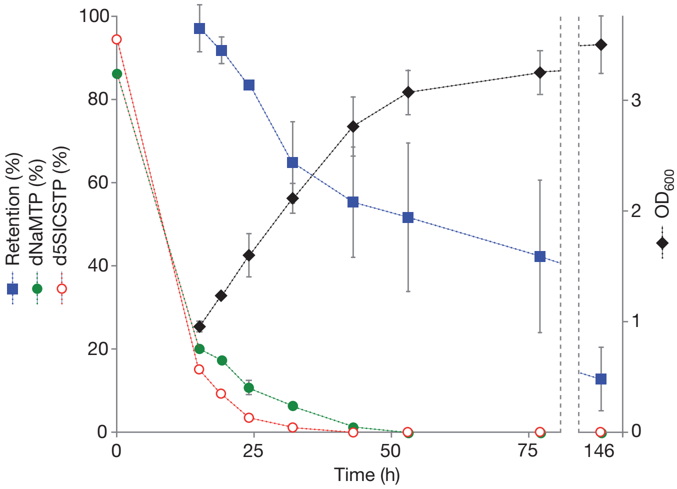


Figure 3 | Intracellular stability of the UBP. *E. coli* C41(DE3) pACS was transformed with pINF and grown after a single dose of d5SICSTP and dNaMTP was provided in the media. UBP retention in recovered pINF, OD_{600}, and relative amount of d5SICSTP and dNaMTP in the media (100% = 0.25 mM), were determined as a function of time. Error bars represent s.d. of the mean, $n = 3$.

with replication (Extended Data Fig. 7b, c, d), but even then, retention of the UBP remained at approximately 45% and 15%, at days 3 and 6 respectively. Moreover, when d5SICS-dNaM was lost, it was replaced by dA–dT, which is consistent with the mutational spectrum of DNA pol I[20]. Finally, the shape of the retention versus time curve mirrors that of the growth versus time curve. Taken together, these data suggest that in the absence of unnatural triphosphates, the UBP is eventually lost by replication-mediated mispairing, and not from the activity of DNA repair pathways.

We have demonstrated that *Pt*NTT2 efficiently imports d5SICSTP and dNaMTP into *E. coli* and that an endogenous polymerase, possibly pol I, efficiently uses the unnatural triphosphates to replicate DNA containing the UBP within the cellular environment with reasonable efficiency and fidelity. Moreover, the UBP appears stable during both exponential and stationary phase growth despite the presence of all DNA repair mechanisms. Remarkably, although expression of *Pt*NTT2 results in a somewhat reduced growth rate, neither the unnatural triphosphates nor replication of the UBP results in significant further reduction in growth. The resulting bacterium is the first organism that stably harbours DNA containing three base pairs. In the future, this organism, or a variant with the UBP incorporated at other episomal or chromosomal loci, should provide a synthetic biology platform to orthogonally re-engineer cells, with applications ranging from site-specific labelling of nucleic acids in living cells to the construction of orthogonal transcription networks and eventually the production and evolution of proteins with multiple, different unnatural amino acids.

METHODS SUMMARY

To prepare electrocompetent C41(DE3) pACS cells, freshly transformed *E. coli* C41(DE3) pACS was grown overnight in $2 \times$ YT medium (1.6% tryptone, 1% yeast extract, 0.5% NaCl) supplemented with streptomycin and KPi. After 100-fold dilution into the same medium and outgrowth at 37 °C to $OD_{600} = 0.20$, IPTG was added to induce expression of *Pt*NTT2. After 40 min, cultures were rapidly cooled, washed with sterile water and resuspended in 10% glycerol. An aliquot of electrocompetent cells was mixed with pINF and electroporated. Pre-warmed $2 \times$ YT medium containing streptomycin, IPTG and KPi was added, and an aliquot was diluted 3.3-fold in the same media supplemented with 0.25 mM each of dNaMTP and d5SICSTP. The resulting mixture was allowed to recover at 37 °C with shaking. After recovery, cultures were centrifuged. Spent media was analysed for nucleotide composition by HPLC (Extended Data Fig. 7a); cells were resuspended in fresh medium containing streptomycin, ampicillin, IPTG, KPi and 0.25 mM each of dNaMTP and d5SICSTP, and grown with shaking. At defined time points, OD_{600} was determined and aliquots were removed and centrifuged. Spent media were analysed for nucleotide

composition, and pINF was recovered by spin column purification. UBP retention was characterized by LC-MS/MS, PCR amplification and gel electrophoresis, or sequencing, as described in the Methods.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 27 November 2013; accepted 8 April 2014.

Published online 7 May 2014.

1. Malyshev, D. A. *et al.* Efficient and sequence-independent replication of DNA containing a third base pair establishes a functional six-letter genetic alphabet. *Proc. Natl Acad. Sci. USA* **109,** 12005–12010 (2012).
2. Yang, Z., Chen, F., Alvarado, J. B. & Benner, S. A. Amplification, mutation, and sequencing of a six-letter synthetic genetic system. *J. Am. Chem. Soc.* **133,** 15105–15112 (2011).
3. Yamashige, R. *et al.* Highly specific unnatural base pair systems as a third base pair for PCR amplification. *Nucleic Acids Res.* **40,** 2793–2806 (2012).
4. Seo, Y. J., Malyshev, D. A., Lavergne, T., Ordoukhanian, P. & Romesberg, F. E. Site-specific labeling of DNA and RNA using an efficiently replicated and transcribed class of unnatural base pairs. *J. Am. Chem. Soc.* **133,** 19878–19888 (2011).
5. Seo, Y. J., Matsuda, S. & Romesberg, F. E. Transcription of an expanded genetic alphabet. *J. Am. Chem. Soc.* **131,** 5046–5047 (2009).
6. Betz, K. *et al.* Structural insights into DNA replication without hydrogen bonds. *J. Am. Chem. Soc.* **135,** 18637–18643 (2013).
7. Betz, K. *et al.* KlenTaq polymerase replicates unnatural base pairs by inducing a Watson–Crick geometry. *Nature Chem. Biol.* **8,** 612–614 (2012).
8. Wu, Y., Fa, M., Tae, E. L., Schultz, P. G. & Romesberg, F. E. Enzymatic phosphorylation of unnatural nucleosides. *J. Am. Chem. Soc.* **124,** 14626–14630 (2002).
9. Yan, H. & Tsai, M. D. Nucleoside monophosphate kinases: structure, mechanism, and substrate specificity. *Adv. Enzymol.* **73,** 103–134 (1999).
10. Winkler, H. H. & Neuhaus, H. E. Non-mitochondrial ATP transport. *Trends Biochem. Sci.* **24,** 64–68 (1999).
11. Amiri, H., Karlberg, O. & Andersson, S. G. Deep origin of plastid/parasite ATP/ADP translocases. *J. Mol. Evol.* **56,** 137–150 (2003).
12. Hatch, T. P., Al-Hossainy, E. & Silverman, J. A. Adenine nucleotide and lysine transport in *Chlamydia psittaci. J. Bacteriol.* **150,** 662–670 (1982).
13. Winkler, H. H. Rickettsial permeability: an ADP-ATP transport system. *J. Biol. Chem.* **251,** 389–396 (1976).
14. Horn, M. & Wagner, M. Bacterial endosymbionts of free-living amoebae. *J. Eukaryot. Microbiol.* **51,** 509–514 (2004).
15. Haferkamp, I. *et al.* Tapping the nucleotide pool of the host: novel nucleotide carrier proteins of *Protochlamydia amoebophila. Mol. Microbiol.* **60,** 1534–1545 (2006).
16. Miroux, B. & Walker, J. E. Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J. Mol. Biol.* **260,** 289–298 (1996).
17. Haferkamp, I. & Linka, N. Functional expression and characterisation of membrane transport proteins. *Plant Biol.* **14,** 675–690 (2012).
18. Ast, M. *et al.* Diatom plastids depend on nucleotide import from the cytosol. *Proc. Natl Acad. Sci. USA* **106,** 3621–3626 (2009).
19. Baba, T. *et al.* Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2,** 2006.0008 (2006).
20. Lavergne, T., Malyshev, D. A. & Romesberg, F. E. Major groove substituents and polymerase recognition of a class of predominantly hydrophobic unnatural base pairs. *Chemistry* **18,** 1231–1239 (2012).
21. Seo, Y. J., Hwang, G. T., Ordoukhanian, P. & Romesberg, F. E. Optimization of an unnatural base pair toward natural-like replication. *J. Am. Chem. Soc.* **131,** 3246–3252 (2009).
22. Li, L. *et al.* Natural-like replication of an unnatural base pair for the expansion of the genetic alphabet and biotechnology applications. *J. Am. Chem. Soc.* **136,** 826–829 (2014).
23. Tomizawa, J. & Selzer, G. Initiation of DNA synthesis in *Escherichia coli. Annu. Rev. Biochem.* **48,** 999–1034 (1979).
24. Allen, J. M. *et al.* Roles of DNA polymerase I in leading and lagging-strand replication defined by a high-resolution mutation footprint of ColE1 plasmid replication. *Nucleic Acids Res.* **39,** 7020–7033 (2011).
25. Hashimoto, H. *et al.* Structure of a *Naegleria* Tet-like dioxygenase in complex with 5-methylcytosine DNA. *Nature* **506,** 391–395 (2013).
26. Malyshev, D. A., Seo, Y. J., Ordoukhanian, P. & Romesberg, F. E. PCR with an expanded genetic alphabet. *J. Am. Chem. Soc.* **131,** 14620–14621 (2009).
27. Hirao, I. *et al.* An unnatural hydrophobic base pair system: site-specific incorporation of nucleotide analogs into DNA and RNA. *Nature Methods* **3,** 729–735 (2006).
28. Goodman, M. F. Error-prone repair DNA polymerases in prokaryotes and eukaryotes. *Annu. Rev. Biochem.* **71,** 17–50 (2002).
29. Quan, J. & Tian, J. Circular polymerase extension cloning for high-throughput cloning of complex and combinatorial DNA libraries. *Nature Protocols* **6,** 242–251 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank I. Haferkamp and J. Audia for kindly providing the NTT plasmids and helpful discussions, and P. Ordoukhanian for providing access to the Center for Protein and Nucleic Acid Research at TSRI. This work was supported by the US National Institutes of Health (NIH) (GM 060005).

Author Contributions D.A.M., K.D., T.C. and F.E.R. designed the experiments. D.A.M., K.D. and T.L. performed the experiments. N.D., J.M.F. and I.R.C.J. performed LC-MS/MS analysis. D.A.M., K.D. and F.E.R. analysed data and D.A.M. and F.E.R. wrote the manuscript with assistance from the other authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to F.E.R. (floyd@scripps.edu).

METHODS

Materials. 2 × YT, 2 × YT agar, IPTG, ampicillin and streptomycin were obtained from Fisher Scientific. Ampicillin and streptomycin were used at 100 μg ml^{-1} and 50 μg ml^{-1}, respectively. All pET-16b constructs containing the nucleotide transporters were kindly provided by I. Haferkamp (Technische Universität Kaiserslautern, Germany) with the exception of pET16b-*Rp*NTT2, which along with the C41(DE3) *E. coli* strain, was provided by J. P. Audia (University of South Alabama, USA). Plasmids pUC19 and pCDF-1b were obtained from Thermo Scientific and EMD Millipore, respectively. Plasmids were purified using the PureLink Quick Plasmid DNA Miniprep Kit (Life Technologies). OneTaq, DeepVent, Q5 Hot Start High-Fidelity DNA Polymerases, and all restriction endonucleases were obtained from New England Biolabs. In general, PCR reactions were divided into multiple aliquots with one followed in real time using 0.5 × Sybr Green I (Life Technologies); following PCR, the aliquots were recombined, purified by spin column (DNA Clean and Concentrator-5; Zymo Research, Irvine, California, USA) with elution in 20 μl of water, then separated by agarose gel electrophoresis, followed by band excision and recovery (Zymoclean Gel DNA Recovery Kit), eluting with 20 μl of water unless stated otherwise. Polyacrylamide gels were stained with 1 × Sybr Gold (Life Technologies) for 30 min, agarose gels were cast with 1 × Sybr Gold. All gels were visualized using a Molecular Imager Gel Doc XR+ equipped with 520DF30 filter (Bio-Rad) and quantified with Quantity One software (Bio-Rad). The sequences of all DNA oligonucleotides used in this study are provided in Supplementary Information. Natural oligonucleotides were purchased from IDT (San Diego, California, USA). The concentration of dsDNA was measured by fluorescent dye binding (Quant-iT dsDNA HS Assay kit, Life Technologies) unless stated otherwise. The concentration of ssDNA was determined by UV absorption at 260 nm using a NanoDrop 1000 (Thermo Scientific). [α-^{32}P]-dATP (25 μCi) was purchased from PerkinElmer (Shelton, Connecticut, USA). Polyethyleneimine cellulose pre-coated Bakerflex TLC plates (0.5 mm) were purchased from VWR. dNaM phosphoramidite, dNaM and d5SICS nucleosides were obtained from Berry & Associates Inc. (Dexter, Michigan, USA). Free nucleosides of dNaM and d5SICS (Berry & Associates) were converted to the corresponding triphosphates under Ludwig conditions[30]. After purification by anion exchange chromatography (DEAE Sephadex A-25) followed by reverse phase (C18) HPLC and elution through a Dowex 50WX2-sodium column, both triphosphates were lyophilized and kept at −20 °C until use. The d5SICSTP analogue dTPT3TP[22] and the biotinylated dNaMTP analogue dmmo2$^{\mathrm{SSBIO}}$TP[4] were made as reported previously. MALDI-TOF mass spectrometry (Applied Biosystems Voyager DE-PRO System 6008) was performed at the TSRI Center for Protein and Nucleic Acid Research.

Construction of NTT expression plasmids. The Pt*NTT2* gene was amplified from plasmid pET-16b-*Pt*NTT2 using primers PtNTT2-forward and PtNTT2-reverse; the TpNTT2 gene was amplified from plasmid pET-16b-*Tp*NTT2 using primers TpNTT2-forward and TpNTT2-reverse. A linear fragment of pCDF-1b was generated using primers pCDF-1b-forward and pCDF-1b-reverse. All fragments were purified as described in Materials. The pCDF-1b fragment (100 ng, 4.4 × 10^{-14} mol) and either the Pt*NTT2* (78 ng, 4.4 × 10^{-14} mol) or Tp*NTT2* (85 ng, 4.4 × 10^{-14} mol) fragment were then assembled together using restriction-free circular polymerase extension cloning[29] in 1 × OneTaq reaction buffer, $MgSO_4$ adjusted to 3.0 mM, 0.2 mM of dNTP, and 0.02 U μl^{-1} of OneTaq DNA under the following thermal cycling conditions: initial denaturation (96 °C, 1 min); 10 cycles of denaturation (96 °C, 30 s), touchdown annealing (54 °C to 49.5 °C for 30 s (−0.5 °C per cycle)), extension of 68 °C for 5 min, and final extension (68 °C, 5 min). Upon completion, the samples were purified and used for heat-shock transformation of *E. coli* XL10. Individual colonies were selected on lysogeny broth (LB)-agar containing streptomycin, and assayed by colony PCR with primers PtNTT2-forward/reverse or TpNTT2-forward/reverse. The presence of the *NTT* genes was confirmed by sequencing and double digestion with ApaI/EcoO109I restriction endonucleases with the following expected pattern: pCDF-1b-*Pt*NTT2 (2,546/2,605 bp), pCDF-1b-*Tp*NTT2 (2,717/2,605 bp), pCDF-1b (1,016/2,605 bp). The complete nucleotide sequence of the pCDF-1b-*Pt*NTT2 plasmid (pACS) is provided in Supplementary Information.

Growth conditions to quantify nucleoside triphosphate uptake. *E. coli* C41(DE3)[16] freshly transformed with pCDF-1b-*Pt*NTT2 was grown in 2 × YT with streptomycin overnight, then diluted (1:100) into fresh 2 × YT medium (1 ml of culture per uptake with [α-^{32}P]-dATP; 2 ml of culture per uptake with d5SICSTP or dNaMTP) supplemented with 50 mM potassium phosphate (KPi) and streptomycin. A negative control with the inactive transporter pET-16b-*Rp*NTT2, was treated identically except ampicillin was used instead of streptomycin. Cells were grown to an OD_{600} of approximately 0.6 and the NTT expression was induced by the addition of IPTG (1 mM). The culture was allowed to grow for another hour (final OD_{600} approximately 1.2) and then assayed directly for uptake as described below using a method adapted from a previous paper[15].

Preparation of media fraction for unnatural nucleoside triphosphate analysis. The experiment was initiated by the addition of either dNaMTP or d5SICSTP (10 mM each) directly to the media to a final concentration of 0.25 mM. Cells were incubated with the substrate with shaking at 37 °C for 30 min and then pelleted (8,000 r.c.f. (relative centrifugal force) for 5 min, 4 °C). An aliquot of the media fraction (40 μl) was mixed with acetonitrile (80 μl) to precipitate proteins[31], and then incubated at 22 °C for 30 min. Samples were either analysed immediately by HPLC or stored at −80 °C until analysis. Analysis began with centrifugation (12,000 r.c.f. for 10 min at 22 °C), then the pellet was discarded, and the supernatant was reduced to approximately 20 μl by SpeedVac, resuspended in buffer A (see below) to a final volume of 50 μl, and analysed by HPLC (see below).

Preparation of cytoplasmic fraction for nucleoside triphosphate analysis. To analyse the intracellular desphosphorylation of the unnatural nucleoside triphosphate, cell pellets were subjected to 3 × 100 μl washes of ice-cold KPi (50 mM). Pellets were then resuspended in 250 μl of ice cold KPi (50 mM) and lysed with 250 μl of lysis buffer L7 of the PureLink Quick Plasmid DNA Miniprep Kit (200 mM NaOH, 1% w/v SDS), after which the resulting solution was incubated at 22 °C for 5 min. Precipitation buffer N4 (350 μl, 3.1 M potassium acetate, pH 5.5) was added, and the sample was mixed to homogeneity. Following centrifugation (>12,000 r.c.f. for 10 min, at 22 °C) the supernatant containing the unnatural nucleotides was applied to a Hypersep C18 solid phase extraction column (Thermo Scientific) prewashed with acetonitrile (1 ml) and buffer A (1 ml, see HPLC protocol for buffer composition). The column was then washed with buffer A and nucleotides were eluted with 1 ml of 50% acetonitrile:50% triethylammonium bicarbonate (TEAB) 0.1 M (pH 7.5). The eluent was reduced to approximately 50 μl in a SpeedVac and its volume was adjusted to 100 μl with buffer A before HPLC analysis.

HPLC protocol and nucleoside triphosphate quantification. Samples were applied to a Phenomenex Jupiter LC column (3 μm C18 300 Å, 250 × 4.6 mm) and subjected to a linear gradient of 0–40% B over 40 min at a flow rate of 1 ml min^{-1}. Buffer A: 95% 0.1 M TEAB, pH 7.5; 5% acetonitrile. Buffer B: 20% 0.1 M TEAB, pH 7.5; 80% acetonitrile. Absorption was monitored at 230, 273, 288, 326 and 365 nm.

Each injection series included two extra control samples containing 5 nmol of dNaMTP or d5SICSTP. The areas under the peaks that corresponded to triphosphate, diphosphate, monophosphate and free nucleoside (confirmed by MALDI-TOF) were integrated for both the control and the unknown samples (described above). After peak integration, the ratio of the unknown peak to the control peak adjusted for the loss from the extraction step (62% and 70% loss for dNaM and d5SICS, respectively, Extended Data Table 3), provided a measure of the amount of each of the moieties in the sample. To determine the relative concentrations of unnatural nucleotide inside the cell, the amount of imported unnatural nucleotide (d**X**TP, μmol) was then divided by the volume of cells, which was calculated as the product of the volume of a single *E. coli* cell (1 μm^3 based on a reported average value[32]; that is, 1 × 10^{-9} μl per cell) and the number of cells in each culture (OD_{600} of 1.0 equal to 1 × 10^9 cells per ml (ref. 32)). The *Rp*NTT2 sample was used as a negative control and its signal was subtracted to account for incomplete washing of nucleotide species from the media.

dATP uptake. To analyse the intracellular desphosphorylation of dATP, after induction of the transporter, the uptake reaction was initiated by the addition of dATP (spiked with [α-^{32}P]-dATP) to a final concentration of 0.25 mM, followed by incubation at 37 °C with shaking for 30 min. The culture was then centrifuged (8,000 r.c.f. for 5 min at 22 °C). Supernatant was analysed by TLC. Cell pellets were washed three times with ice-cold KPi (50 mM, 100 μl) to remove excess radioactive substrate, lysed with NaOH (0.2 M, 100 μl) and centrifuged (10,000 r.c.f. for 5 min at 22 °C) to remove cell debris; supernatant was analysed by TLC.

TLC analysis. Samples (1 μl) were applied on a 0.5 mm polyethyleneimine cellulose TLC plate and developed with sodium formate pH 3.0 (0.5 M, 30 s; 2.5 M, 2.5 min; 4.0 M, 40 min). Plates were dried using a heat gun and quantified by phosphorimaging (Storm Imager, Molecular Dynamics) and Quantity One software.

Optimization of nucleotide extraction from cells for HPLC injection. To minimize the effect of the lysis and triphosphate extraction protocols on the decomposition of nucleoside triphosphate within the cell, the extraction procedure was optimized for the highest recovery with the lowest extent of decomposition (Extended Data Table 3). To test different extraction methods, cells were grown as described above, washed, and then 5 nmol of either dNaMTP or d5SICSTP was added to the pellets, which were then subjected to different extraction protocols including boiling water, hot ethanol, cold methanol, freeze and thaw, lysozyme, glass beads, NaOH, trichloroacetic acid (TCA) with Freon, and perchloric acid (PCA) with KOH[33]. The recovery and composition of the control was quantified by HPLC as described above to determine the most effective procedure. Method 3—that is, cell lysis with NaOH (Extended Data Table 3)—was found to be most effective and reproducible, thus we further optimized it by resuspension of the pellets in ice-cold KPi (50 mM, 250 μl) before addition of NaOH to decrease dephosphorylation after cell lysis (Method 4). Cell pellets were then processed as described above. See above for the final extraction protocol.

Preparation of the unnatural insert for pINF construction. The TK-1-dNaM oligonucleotide containing dNaM was prepared using solid-phase DNA synthesis

with ultra-mild DNA synthesis phosphoramidites on CPG ultramild supports (1 μmol, Glen Research, Sterling, Virginia, USA) and an ABI Expedite 8905 synthesizer. After the synthesis, the DMT-ON oligonucleotide was cleaved from the solid support, deprotected and purified by Glen-Pak cartridge according to the manufacturer's recommendation (Glen Research), and then subjected to 8 M urea 8% PAGE. The gel was visualized by ultraviolet shadowing, the band corresponding to the 75-mer was excised, and the DNA was recovered by crush and soak extraction, filtration (0.45 μm), and final desalting over Sephadex G-25 (NAP-25 Columns, GE Healthcare). The concentration of the single stranded oligonucleotide was determined by ultraviolet absorption at 260 nm assuming that the extinction coefficient of dNaM at 260 nm is equal to that of dA. TK-1-dNaM (4 ng) was next amplified by PCR under the following conditions: 1 × OneTaq reaction buffer, $MgSO_4$ adjusted to 3.0 mM, 0.2 mM of dNTP, 0.1 mM of dNaMTP, 0.1 mM of the d5SICSTP analogue dTPT3TP, 1 μM of each of the primers pUC19-fusion-forward and pUC19-fusion-reverse, and 0.02 U μl^{-1} of OneTaq DNA Polymerase (in a total of 4 × 50 μl reactions) under the following thermal cycling conditions: initial denaturation (96 °C, 1 min) followed by 12 cycles of denaturation (96 °C, 10 s), annealing (60 °C, 15 s), and extension (68 °C, 2 min). An identical PCR without the unnatural triphosphates was run to obtain fully natural insert under identical conditions for the construction of the natural control plasmid. Reactions were subjected to spin column purification and then the desired PCR product (122 bp) was purified by a 4% agarose gel.

pUC19 linearization for pINF construction. pUC19 (20 ng) was amplified by PCR under the following conditions: 1 × Q5 reaction buffer, $MgSO_4$ adjusted to 3.0 mM, 0.2 mM of dNTP, 1 μM of each primers pUC19-lin-forward and pUC19-lin-reverse, and 0.02 U μl^{-1} of Q5 Hot Start High-Fidelity DNA Polymerase (in a total of 4 × 50 μl reactions with one reaction containing 0.5 × Sybr Green I) under the following thermal cycling conditions: initial denaturation (98 °C, 30 s); 20 cycles of denaturation (98 °C, 10 s), annealing (60 °C, 15 s), and extension (72 °C, 2 min); and final extension (72 °C, 5 min). The desired PCR product (2,611 bp) was purified by a 2% agarose gel.

PCR assembly of pINF and the natural control plasmid. A linear fragment was amplified from pUC19 using primers pUC19-lin-forward and pUC19-lin-reverse. The resulting product (800 ng, 4.6×10^{-13} mol) was combined with either the natural or unnatural insert (see above) (56 ng, 7.0×10^{-13} mol) and assembled by circular overlap extension PCR under the following conditions: 1 × OneTaq reaction buffer, $MgSO_4$ adjusted to 3.0 mM, 0.2 mM of dNTP, 0.1 mM of dNaMTP, 0.1 mM of the d5SICSTP analogue dTPT3TP, and 0.02 U μl^{-1} of OneTaq DNA Polymerase (in a total of 4 × 50 μl reactions with one reaction containing 0.5 × Sybr Green I) using the following thermal cycling conditions: initial denaturation (96 °C, 1 min); 12 cycles of denaturation (96 °C, 30 s), annealing (62 °C, 1 min), and extension (68 °C, 5 min); final extension (68 °C, 5 min); and slow cooling (68 °C to 10 °C at a rate of -0.1 °C s^{-1}). The PCR product was analysed by restriction digestion on 1% agarose and used directly for *E. coli* transformation. The d5SICS analogue dTPT3[22] pairs with dNaM, and dTPT3TP was used in place of d5SICSTP as DNA containing dTPT3–dNaM is better PCR amplified than DNA containing d5SICS–dNaM, and this allowed for differentiation of synthetic and *in vivo* replicated pINF, as well as facilitated the construction of high-quality pINF (UBP content >99%).

Preparation of electrocompetent cells for pINF replication in *E. coli*. C41(DE3) cells were transformed by heat shock[34] with 200 ng of pACS plasmid, and the transformants were selected overnight on 2 × YT-agar supplemented with streptomycin. A single clone of freshly transformed C41(DE3) pACS was grown overnight in 2 × YT medium (3 ml) supplemented with streptomycin and KPi (50 mM). After 100-fold dilution into the same fresh 2 × YT media (300 ml), the cells were grown at 37 °C until they reached an OD_{600} of 0.20 at which time IPTG was added to a final concentration of 1 mM to induce the expression of *Pt*NTT2. Cells were grown for another 40 min and then growth was stopped by rapid cooling in ice water with intensive shaking. After centrifugation in a prechilled centrifuge (2,400 r.c.f. for 10 min, 4 °C), the spent media was removed, and the cells were prepared for electroporation by washing with ice-cold sterile water (3 × 150 ml). After washing, the cells were resuspended in ice-cold 10% glycerol (1.5 ml) and split into 50-μl aliquots. Although we found that dry ice yielded better results than liquid nitrogen for freezing cells to store for later use, freshly prepared cells were used for all reported experiments as they provided higher transformation efficiency of pINF and higher replication fidelity of the UBP.

Electroporation and recovery for pINF replication in *E. coli*. The aliquot of cells was mixed with 2 μl of plasmid (400 ng), transferred to 0.2 cm gap electroporation cuvette and electroporated using a Bio-Rad Gene Pulser according to the manufacturer's recommendations (voltage 25 kV, capacitor 2.5 μF, resistor 200 Ω, time constant 4.8 ms). Pre-warmed 2 × YT media (0.95 ml, streptomycin, 1 mM IPTG, 50 mM KPi) was added, and after mixing, 45 μl was removed and combined with 105 μl of the same media (3.33-fold dilution) supplemented with 0.25 mM of dNaMTP and d5SICSTP. The resulting mixture was allowed to recover for 1 h at 37 °C with shaking (210 revolutions per min (r.p.m.)). The original transformation media (10 μl) was spread onto 2 × YT-agar containing streptomycin with 10- and 50-fold dilutions for the determination of viable colony forming units after overnight growth at 37 °C to calculate the number of the transformed pINF molecules (see the section on calculation of the plasmid amplification). Transformation, recovery and growth were carried out identically for the natural control plasmid. In addition, a negative control was run and treated identically to pINF transformation except that it was not subjected to electroporation (Extended Data Fig. 7b). No growth in the untransformed negative control samples was observed even after 6 days. No PCR amplification of the negative control was detected, which confirms that unamplified pINF plasmid is not carried through cell growth and later detected erroneously as the propagated plasmid.

Analysis of pINF replication in *E. coli*. After recovery, the cells were centrifuged (4,000 r.c.f. for 5 min, 4 °C), and spent media (0.15 ml) was removed and analysed for nucleotide composition by HPLC (Extended Data Fig. 7a). The cells were resuspended in fresh 2 × YT media (1.5 ml, streptomycin, ampicillin, 1 mM IPTG, 50 mM KPi, 0.25 mM dNaMTP, 0.25 mM d5SICSTP) and grown overnight at 37 °C while shaking (250 r.p.m.), resulting in tenfold dilution compared to recovery media or 33.3-fold dilution compared to the originally transformed cells. Aliquots (100 μl) were taken after 15, 19, 24, 32, 43, 53, 77 and 146 h, OD_{600} was determined, and the cells were centrifuged (8,000 r.c.f. for 5 min, 4 °C). Spent media were analysed for nucleotide composition by HPLC (Extended Data Fig. 7a), and the pINF and pACS plasmid mixtures were recovered and linearized with *NdeI* restriction endonuclease; pINF plasmid was purified by 1% agarose gel electrophoresis (Extended Data Fig. 7b) and analysed by LC-MS/MS. The retention of the UBP on the pINF plasmid was quantified by biotin gel shift mobility assay and sequencing as described below.

Mass spectrometry of pINF. Linearized pINF was digested to nucleosides by treatment with a mixture of nuclease P1 (Sigma-Aldrich), shrimp alkaline phosphatase (NEB), and DNase I (NEB), overnight at 37 °C, following a previously reported protocol[25]. LC-MS/MS analysis was performed in duplicate by injecting 15 ng of digested DNA on an Agilent 1290 UHPLC equipped with a G4212A diode array detector and a 6490A Triple Quadrupole Mass Detector operating in the positive electrospray ionization mode (+ESI). UHPLC was carried out using a Waters XSelect HSS T3 XP column (2.1 × 100 mm, 2.5 μm) with the gradient mobile phase consisting of methanol and 10 mM aqueous ammonium formate (pH 4.4). MS data acquisition was performed in Dynamic Multiple Reaction Monitoring (DMRM) mode. Each nucleoside was identified in the extracted chromatogram associated with its specific MS/MS transition: dA at m/z 252→136, d5SICS at m/z 292→176, and dNaM at m/z 275→171. External calibration curves with known amounts of the natural and unnatural nucleosides were used to calculate the ratios of individual nucleosides within the samples analysed. LC-MS/MS quantification was validated using synthetic oligonucleotides[1] containing unnatural d5SICS and dNaM (Extended Data Table 2).

DNA biotinylation by PCR to measure fidelity by gel shift mobility assay. Purified mixtures of pINF and pACS plasmids (1 ng) from growth experiments were amplified by PCR under the following conditions: 1 × OneTaq reaction buffer, $MgSO_4$ adjusted to 3.0 mM, 0.3 mM of dNTP, 0.1 mM of the biotinylated dNaMTP analogue $dMMO2^{SSBIO}TP$, 0.1 mM of d5SICSTP, 1 μM of each of the primers pUC19-seq-forward and pUC19-seq-reverse, 0.02 U μl^{-1} of OneTaq DNA Polymerase, and 0.0025 U μl^{-1} of DeepVent DNA Polymerase in a total volume of 25 μl in an CFX Connect Real-Time PCR Detection System (Bio-Rad) under the following thermal cycling conditions: initial denaturation (96 °C, 1 min); 10 cycles of denaturation (96 °C, 30 s), annealing (64 °C, 30 s), and extension (68 °C, 4 min). PCR products were purified, and the resulting biotinylated DNA duplexes (5 μl, 25–50 ng) were mixed with streptavidin (1 μl, 1 μg μl^{-1}, Promega) in phosphate buffer (50 mM sodium phosphate, pH 7.5, 150 mM NaCl, 1 mM EDTA), incubated for 30 min at 37 °C, mixed with 5 × non-denaturing loading buffer (Qiagen), and loaded onto 6% non-denaturing PAGE. After running at 110 V for 30 min, the gel was visualized and quantified. The resulting fragment (194 bp) with primer regions underlined and the unnatural nucleotide in bold (X represents dNaM or its biotinylated analogue $dMMO2^{SSBIO}$) is 5′-GCAGGCATGCAAGCTTGGCGTAATCATGG TCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACA**X**TTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAG.

Streptavidin shift calibration for gel shift mobility assay. We have already reported a calibration between streptavidin shift and the fraction of sequences with UBP in the population (see Supplementary Fig. 8 of ref. 1). However, we found that spiking the PCR reaction with DeepVent improves the fidelity with which DNA containing d5SICS-$dMMO2^{SSBIO}$ is amplified, and thus we repeated the calibration with added DeepVent. To quantify the net retention of the UBP, nine defined mixtures of the TK-1-dNaM template and its fully natural counterpart were prepared (Extended Data Fig. 6a), subjected to biotinylation by PCR and analysed by mobility-shift assay on 6% non-denaturing PAGE as described above. For calibration, the mixtures

TK-1-dNaM template and its fully natural counterpart with a known ratio of unnatural and natural templates (0.04 ng) were amplified under the same conditions over nine cycles of PCR with pUC19-fusion primers and analysed identically to samples from the growth experiment (see the section on DNA biotinylation by PCR). Each experiment was run in triplicate (a representative gel assay is shown in Extended Data Fig. 6b), and the streptavidin shift (SAS, %) was plotted as function of the UBP content (UBP, %). The data was then fit to a linear equation, $SAS = 0.77 \times UBP + 2.0$ ($R^2 = 0.999$), where UBP corresponds to the retention of the UBP (%) in the analysed samples after cellular replication and was calculated from the SAS shift using the equation above.

Calculation of plasmid amplification. The cells were plated on $2\times$ YT-agar containing ampicillin and streptavidin directly after transformation with pINF, and the colonies were counted after overnight growth at 37 °C. Assuming each cell is only transformed with one molecule of plasmid, colony counts correspond to the original amount of plasmid that was taken up by the cells. After overnight growth, the plasmids were purified from a specific volume of the cell culture and quantified. As purified plasmid DNA represents a mixture of the pINF and pACS plasmids, digestion restriction analysis with NdeI exonuclease was performed to linearize both plasmids, followed by 1% agarose gel electrophoresis (Extended Data Fig. 7b). An example of calculations for the 19-h time point with one of three triplicates is provided in Supplementary Information.

Fragment generation for Sanger sequencing to measure fidelity. Purified mixtures of pINF and pACS plasmids (1 ng) after the overnight growth were amplified by PCR under the following conditions: $1\times$ OneTaq reaction buffer, $MgSO_4$ adjusted to 3.0 mM, 0.2 mM of dNTP, 0.1 mM of dNaMTP, 0.1 mM of the d5SICSTP analogue dTPT3TP, 1 μM of each of the primers pUC19-seq2-forward and pUC19-seq-reverse (see below), and 0.02 U μl^{-1} of OneTaq DNA Polymerase in a total volume of 25 μl under the following thermal cycling conditions: initial denaturation (96 °C, 1 min); and 10 cycles of denaturation (96 °C, 30 s), annealing (64 °C, 30 s), and extension (68 °C, 2 min). Products were purified by spin column, quantified to measure DNA concentration and then sequenced as described below. The sequenced fragment (304 bp) with primer regions underlined and the unnatural nucleotide in bold (X, dNaM) is 5′-GCTGCAAGGCGATTAAGTTGGGTAACGCC AGGGT TTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAATTCGAGCTCG GTACCCGGGGATCCTCTAGAGTCGACCTGCAGGCATGCAAGCTTGGCG TAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACA**X** TTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTG CCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCG CTTTCCAGTCGGGAAACCTGTCGTGCCAG.

Sanger sequencing. The cycle sequencing reactions (10 μl) were performed on a 9800 Fast Thermal Cycler (Applied Biosystems) with the Cycle Sequencing Mix (0.5 μl) of the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) containing 1 ng template and 6 pmol of sequencing primer pUC19-seq-reverse under the following thermal cycling conditions: initial denaturation (98 °C, 1 min); and 25 cycles of denaturation (96 °C, 10 s), annealing (60 °C, 15 s), and extension (68 °C, 2.5 min). Upon completion, the residual dye terminators were removed from the reaction with Agencourt CleanSEQ (Beckman-Coulter, Danvers, Massachusetts, USA). Products were eluted off the beads with deionized water and sequenced directly on a 3730 DNA Analyzer (Applied Biosystems). Sequencing traces were collected using Applied Biosystems Data Collection software v3.0 and analysed with the Applied Biosystems Sequencing Analysis v5.2 software.

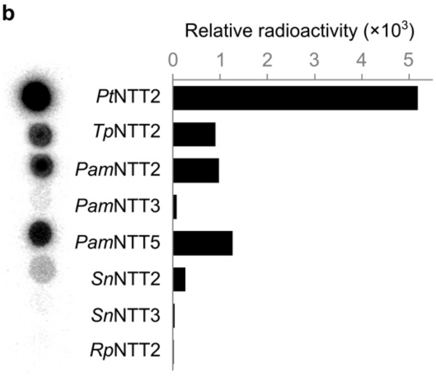
Analysis of Sanger sequencing traces. Sanger sequencing traces were analysed as described previously[1,26] to determine the retention of the unnatural base pair. In brief, the presence of an unnatural nucleotide leads to a sharp termination of the sequencing profile, whereas mutation to a natural nucleotide results in 'read-through'. The extent of this read-through after normalization is inversely correlated with the retention of the unnatural base pair. Raw sequencing traces were analysed by first adjusting the start and stop points for the Sequencing Analysis software (Applied Biosystems) and then determining the average signal intensity individually for each channel (A, C, G and T) for peaks within the defined points. This was done separately for the parts of the sequencing trace before (section L) and after (section R) the unnatural nucleotide. The R/L ratio after normalization $(R/L)_{norm}$ for sequencing decay and read-through in the control unamplified sample ($R/L = 0.55(R/L)_{norm} + 7.2$, see ref. 26 for details) corresponds to the percentage of the natural sequences in the pool. Therefore, an overall retention (F) of the incorporation of the unnatural base pair during PCR is equal to $1 - (R/L)_{norm}$. As significant read-through (over 20%) was observed in the direction of the pUC19-seq2-forward primer even with the control plasmid (synthetic pINF); sequencing of only the opposite direction (pUC19-seq-reverse) was used to gauge fidelity. Raw sequencing traces are shown in Fig. 2f and provided as Supplementary Data.

30. Ludwig, J. & Eckstein, F. Rapid and efficient synthesis of nucleoside 5′-0-(1-thiotriphosphates), 5′-triphosphates and 2′,3′-cyclophosphorothioates using 2-chloro-4H-1,3,2-benzodioxaphosphorin-4-one. *J. Org. Chem.* **54,** 631–635 (1989).
31. Alpert, A. & Shukla, A. Precipitation of Large, High-Abundance Proteins from Serum with Organic Solvents in *ABRF 2003: Translating Biology using Proteomics and Functional Genomics* Poster no. P111-W http://www.abrf.org/Other/ABRFMeetings/ABRF2003/Alpert.pdf (2003).
32. Kubitschek, H. E. & Friske, J. A. Determination of bacterial cell volume with the Coulter Counter. *J. Bacteriol.* **168,** 1466–1467 (1986).
33. Yanes, O., Tautenhahn, R., Patti, G. J. & Siuzdak, G. Expanding coverage of the metabolome for global metabolite profiling. *Anal. Chem.* **83,** 2152–2161 (2011).
34. Seidman, C. E., Struhl, K., Sheen, J. & Jessen, T. Introduction of plasmid DNA into cells. *Curr. Prot. Mol. Biol.* Chapter 1, Unit 1.8 (2001).
35. Knab, S., Mushak, T. M., Schmitz-Esser, S., Horn, M. & Haferkamp, I. Nucleotide parasitism by *Simkania negevensis* (Chlamydiae). *J. Bacteriol.* **193,** 225–235 (2011).
36. Audia, J. P. & Winkler, H. H. Study of the five *Rickettsia prowazekii* proteins annotated as ATP/ADP translocases (Tlc): Only Tlc1 transports ATP/ADP, while Tlc4 and Tlc5 transport other ribonucleotides. *J. Bacteriol.* **188,** 6261–6268 (2006).
37. Hofer, A., Ekanem, J. T. & Thelander, L. Allosteric regulation of *Trypanosoma brucei* ribonucleotide reductase studied *in vitro* and *in vivo*. *J. Biol. Chem.* **273,** 34098–34104 (1998).
38. Reijenga, J. C., Wes, J. H. & van Dongen, C. A. M. Comparison of methanol and perchloric acid extraction procedures for analysis of nucleotides by isotachophoresis. *J. Chromatogr. B Biomed. Sci. Appl.* **374,** 162–169 (1986).

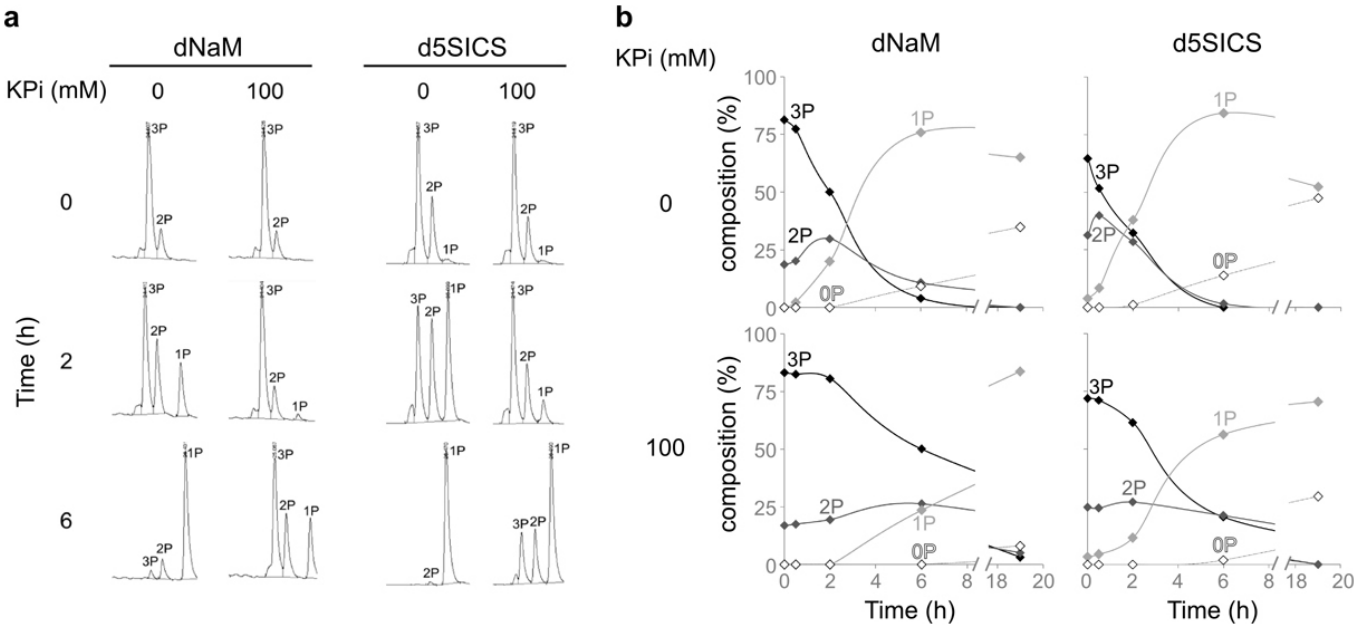
a

Transporter	Deoxyribotriphosphates				Ribotriphosphates				Reference
	dG	dA	dC	T	G	A	C	U	
*Pt*NTT2	82	271	31	428	78	197	49	86	18
*Tp*NTT2	74	665	60	251	59	49	33	90	18
*Pam*NTT2					156	437	570	676	15
*Pam*NTT3								1320	15
*Pam*NTT5	121				22	360			15
*Sn*NTT2					179	654			35
*Sn*NTT3			42		407	375	9	34	35
*Rp*NTT2	substrate is unknown								36

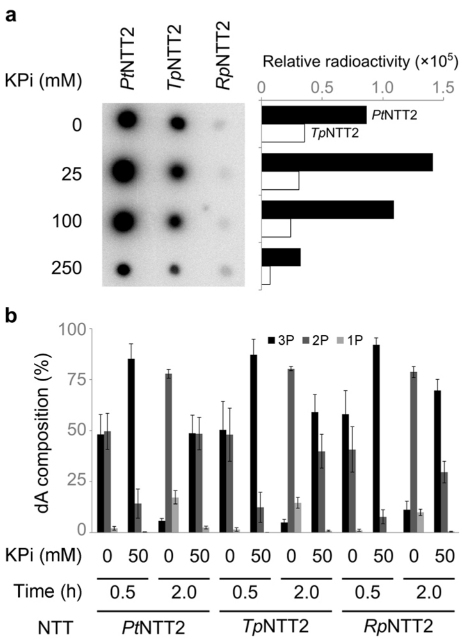
b



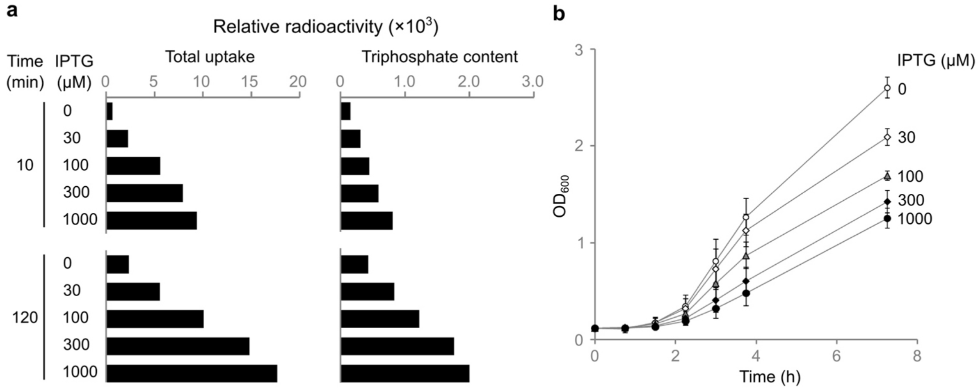
Extended Data Figure 1 | Natural triphosphate uptake by NTTs. a, Survey of reported substrate specificity (K_M, μM) of the NTTs assayed in this study. **b**, *Pt*NTT2 is significantly more active in the uptake of [α-^{32}P]-dATP compared to other nucleotide transporters. Raw (left) and processed (right) data are shown. Relative radioactivity corresponds to the total number of counts produced by each sample. Interestingly, both *Pam*NTT2 and *Pam*NTT5 exhibit a measurable uptake of dATP although this activity was not reported before. This can possibly be explained by the fact that substrate specificity was only characterized using competition experiments, and assay sensitivity might not have been adequate to detect this activity[15]. References 35, 36 are cited in this figure.



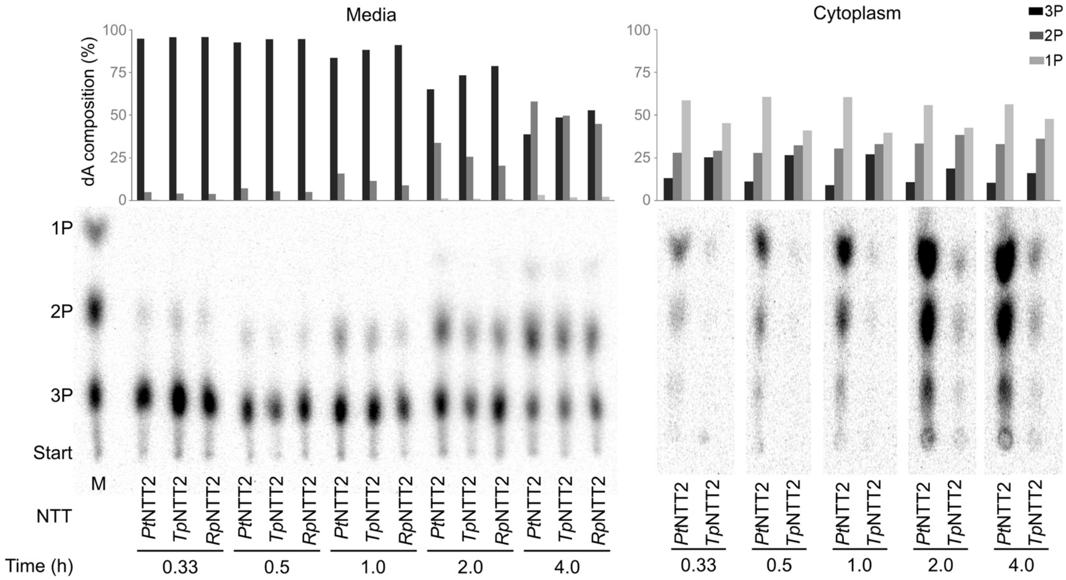
Extended Data Figure 2 | Degradation of unnatural triphosphates in growth media. Unnatural triphosphates (3P) of dNaM and d5SICS are degraded to diphosphates (2P), monophosphates (1P) and nucleosides (0P) in the growing bacterial culture. Potassium phosphate (KPi) significantly slows down the dephosphorylation of both unnatural triphosphates. **a**, Representative HPLC traces (for the region between ~20 and 24 min). dNaM and d5SICS nucleosides are eluted at approximately 40 min and not shown. **b**, Composition profiles.



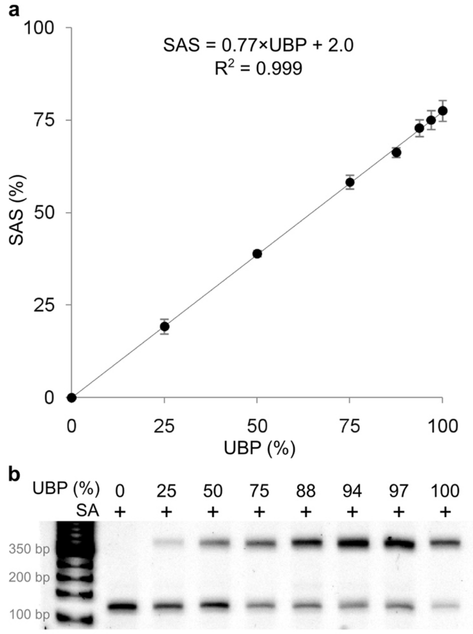
Extended Data Figure 3 | Effect of potassium phosphate on dATP uptake and stability in growth media. **a**, KPi inhibits the uptake of [α-^{32}P]-dATP at concentrations above 100 mM. Raw (left) and processed (right) data are shown. The NTT from *Rickettsia prowazekii* (*Rp*NTT2) does not mediate the uptake of any of the dNTPs and was used as a negative control: its background signal was subtracted from those of *Pt*NTT2 (black bars) and *Tp*NTT2 (white bars). Relative radioactivity corresponds to the total number of counts produced by each sample. **b**, KPi (50 mM) significantly stabilizes [α-^{32}P]-dATP in the media. Triphosphate stability in the media is not significantly affected by the nature of the NTT expressed. 3P, 2P and 1P correspond to triphosphate, diphosphate and monophosphate states, respectively. Error bars represent s.d. of the mean, $n = 3$.



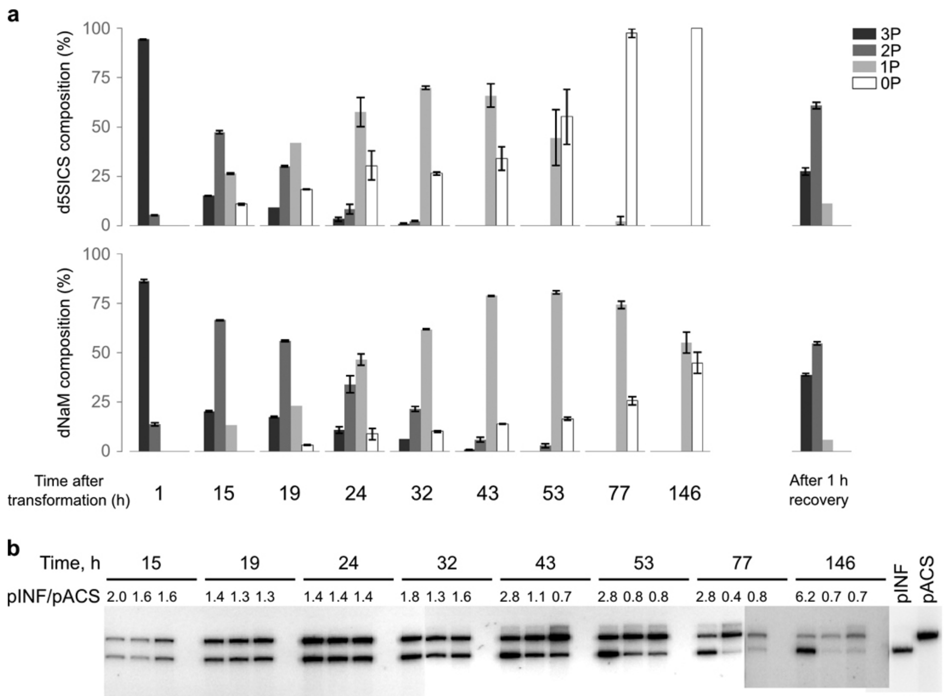
Extended Data Figure 4 | dATP uptake and growth of cells expressing *Pt*NTT2 as a function of inducer (IPTG) concentration. Growth curves and [α-^{32}P]-dATP uptake by bacterial cells transformed with pCDF-1b-*Pt*NTT2 (pACS) plasmid as a function of IPTG concentration. **a**, Total uptake of radioactive substrate (left) and total intracellular triphosphate content (right) are shown at two different time points. Relative radioactivity corresponds to the total number of counts produced by each sample. **b**, A stationary phase culture of C41(DE3) pACS cells was diluted 100-fold into fresh 2 × YT media containing 50 mM KPi, streptomycin, and IPTG at the indicated concentrations and were grown at 37°C. Error bars represent s.d. of the mean, $n = 3$.

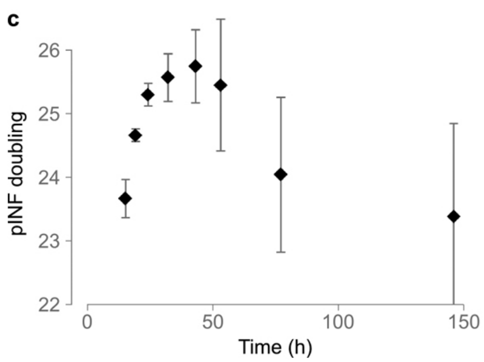


Extended Data Figure 5 | Stability and uptake of dATP in the presence of 50 mM KPi and 1 mM IPTG. Composition of [α-^{32}P]-dATP in the media (left) and cytoplasmic fraction (right) as a function of time. TLC images and their quantifications are shown at the bottom and the top of each of the panels, respectively. 3P, 2P and 1P correspond to nucleoside triphosphate, diphosphate and monophosphate, respectively. M refers to a mixture of all three compounds that was used as a TLC standard. The position labelled 'Start' corresponds to the position of sample spotting on the TLC plate.



Extended Data Figure 6 | Calibration of the streptavidin shift (SAS). **a**, The SAS is plotted as a function of the fraction of template containing the UBP. Error bars represent s.d. of the mean, $n = 3$. **b,** Representative data. SA, streptavidin.





d

Time (h)	Retention (%) Sequencing	Gel shift
15	> 95	97 ± 5
19	> 95	91 ± 3
24	> 95	83 ± 1
32	85 ± 5	64 ± 10
43	71 ± 10	55 ± 13
53	68 ± 21	51 ± 18
77	60 ± 16	42 ± 18
146	30 ± 12	12 ± 8

Extended Data Figure 7 | Decomposition of unnatural triphosphates, pINF quantification, and retention of the UBP with extended cell growth. **a**, Dephosphorylation of the unnatural nucleoside triphosphate. 3P, 2P, 1P and 0P correspond to triphosphate, diphosphate, monophosphate and nucleoside states, respectively. The composition at the end of the 1 h recovery is shown at the right. **b**, Restriction analysis of pINF and pACS plasmids purified from *E. coli*, linearized with NdeI restriction endonuclease and separated on a 1% agarose gel (assembled from independent gel images). Molar ratios of pINF/pACS plasmids are shown at the top of each lane. For each time point, triplicate data are shown in three lanes with the untransformed control shown in the fourth, rightmost lane (see Methods). **c**, Number of pINF doublings as a function of time. The decrease starting at approximately 50 h is due to the loss of the pINF plasmid that also results in increased error. See the section on pINF replication in *E. coli* in the Methods for details. **d**, UBP retention (%) as a function of growth as determined by gel shift (data shown in Fig. 3) and Sanger sequencing (sequencing traces are available as Supplementary Data). In **a**, **c** and **d**, error shown is the s.d. of mean, $n = 3$.

Extended Data Table 1 | OD_{600} of *E. coli* cultures and relative copy number of plasmid (pINF or control pUC19) as determined by its molar ratio to pACS after 19 h of growth

Plasmid	IPTG	dXTP/dYTP	Relative Copy Number	OD_{600} (15 h)	OD_{600} (19 h)
pINF	+	+	1.8	0.34	0.75
	+	–	4.6	0.15	0.75
	–	+	8.9	3.13	3.98
pUC19	+	+	2.8	0.54	1.25
	+	–	2.6	0.73	1.39

X, NaM; Y, 5SICS.

Extended Data Table 2 | Relative quantification by LC-MS/MS using synthetic oligonucleotides containing d5SICS and dNaM

Oligonucleotide (ssDNA)	Size (nt)	Sequence	dA/d5SICS Exp. (calcd.)	dA/dNaM Exp. (calcd.)
D6-NaM	82	CAC ACA GGA AAC AGC TAT GAC CCG GGT TAT TAC ATG CGC TAG CAC TTG GAA TTC ACC AG ACG NNN NaM NNN CGG GAC CCA TAG T		**22.5** (23.5)*
D6-5SICS	87	GAA ATT AAT ACG ACT CAC TAT AGG GTT AAG CTT AAC TTT AAG AAG GAG ATT TAC TAT GGG TCC CG NNN 5SICS NNN CGT CTG GTG AAT TCC	**23.4** (25.5)*	
D13-NaM×2	130	CAC ACA GGA AAC AGC TAT GAC CCG GGT TAT TAC ATG CGC TAG CAC TTG GAA TTC ACT ATC AC NaM AGT CAC NaM AGT AAT CCA TAG TAA ATC TCC TTG TTA AGC TTA ACC CTA TAG TGA GTC GTA TTA ATT TCT		**16.1** (19.5)

* dA/d5SICS and dA/dNaM ratios were calculated assuming that randomized nucleotides (N) around the unnatural base are distributed equally.

Extended Data Table 3 | Summary of the most successful extraction methods

Method		Protocol summary	Total recovery (%)*		Triphosphate stability (%)†		Ref.
			dNaM	d5SICS	dNaM	d5SICS	
1.	TCA with Freon	• Lyse with cold TCA • Extract aqueous phase using Freon with trioctylamine solution	38	23	92	99	Adapted from Ref. 37
2.	PCA w/KOH	• Lyse with cold PCA • Precipitate proteins with KOH and KPi	36	21	98	77	Adapted from Ref. 38
3.	NaOH w/ KOAc	• Lyse with NaOH and SDS • Precipitate proteins with potassium acetate	21	26	86	100	see footnotes
4.	NaOH w/ KOAc supplemented w/ KPi (50 mM)	• Suspend cells in KPi • Lyse with NaOH and SDS • Precipitate proteins with potassium acetate	38	30	99	100	see footnotes

* Recovery of all nucleotides (3P, 2P, 1P and nucleoside).
† Calculated as a ratio of 3P composition (%) before and after the extraction.
References 37, 38 are cited in this figure. Details of methods 3 and 4 can be found online (http://2013.igem.org/wiki/images/e/ed/BGU_purelink_quick_plasmid_qrc.pdf).

TURNING POINT Researcher battles learning disabilities to study HIV **p.391**

@NATUREJOBS Follow us on Twitter for the latest news and features **go.nature.com/e492gf**

NATUREJOBS For the latest career listings and advice **www.naturejobs.com**

AMY DEVOOGD/GETTY

CAREER GAPS

Maternity muddle

The support available for childbirth and rearing varies wildly. New parents come up with creative ways to juggle demands.

BY AMANDA MASCARELLI

Marine scientist Maria Granberg gave birth to her first child in 2006 and took roughly a year off at 80% pay. But in 2010, when she became pregnant with her second child, she was at a more challenging career stage. By then an assistant professor at the University of Gothenburg in Sweden, she was building momentum, planning research activities and establishing collaborations — and had yet to recruit staff for her lab to keep the research moving in her absence.

In addition, her research on marine-sediment bacteria involved work with solvents that are unsafe to handle during pregnancy. Neither her department nor her granting body, the Swedish Research Council Formas, could offer her funding for a lab assistant. Between a pregnancy that restricted her work and six months of partial maternity leave, she lost a full season of field samples and a year of work. The losses resulted in missed publication opportunities and altered her long-term plans for the project. "That's a very stressful thing," she says. "The pregnancy comes along and kind of disturbs your plan."

Four years on, Granberg says that she is finally regaining traction, but she still views that period as troublesome to her career. "Science is a lot about timing, having momentum, and I think the scientific community could do a lot about this period when you are pregnant and you cannot be in the lab," she says. For example, Granberg believes that research councils and granting agencies should establish funding to support female researchers' programmes during pregnancy and directly after their return from a maternity leave. "Efforts should be aimed at sustaining the research momentum at the same level as that of their colleagues," she says.

Her experience illustrates the challenges for women who are trying to juggle the obligations of the lab with the demands of motherhood. The decision to start a family often clashes with the time when researchers are launching their careers and working towards tenure. Maternity-leave benefits and experiences vary widely among employers and at institutions across the globe, and even between departments. The amount of time a woman spends working during her leave depends on variables such as personal desires, career phase, the nature of the research and the institution's policies.

Most researcher parents agree that there is no perfect time in a career for the arrival of a child. Already having tenure might offer job security, but it also comes with the responsibility of overseeing lab members and projects. Having a child during the early days of a career means that there may not be major projects or personnel to oversee — but it can interrupt the momentum of the research.

Early-career researchers who are considering parenthood or who are preparing for a child need to be aware of the challenges that are likely to await them. Depending on their location and employer, some or all of their maternity leave may have to be unpaid, they may need to work from home at times, and they could well need to find ways to keep their research programme afloat.

It is a tough road to navigate, especially for those who are just getting under way. People who have travelled down that road warn that junior researchers who are considering parenthood should find out specifically and well in advance what benefits their employer offers and be ready to negotiate if they do not exist or do not meet their needs.

AVENUES OF SUPPORT

No matter how supportive the work environment, breaking away from research can be fraught with difficult decisions, and there is no universal formula for how to satisfy the twin demands of work and family. Some ▶

institutions and agencies offer support through a mixture of programmes and policies — but with varying levels of success.

The US National Science Foundation in Arlington, Virginia, for example, is starting to incorporate 'family-friendly practices' into its framework. Its Career–Life Balance Initiative was launched in 2011 and provides, among other offerings, supplemental funding of three months' salary — about US$12,000 — to help principal investigators, postdocs and graduate research fellows to pay for a technician during their maternity leave. And the European Research Council has adapted its rules to allow principal investigators who become parents additional time for each child born to apply for specific grants.

TIME TO NEGOTIATE

Some institutions offer very little in the way of formal support. Gretchen Hansen was finishing her PhD in limnology and marine science at the University of Wisconsin–Madison when she became pregnant with her daughter. She told her adviser about her pregnancy and asked about the maternity policy for graduate students. "Nobody knew," she says, "and there was no policy."

She and her adviser then devised a plan that would allow her to drop down to a 33% appointment during her leave, reduce her salary by one-third and keep her health insurance. Hansen spent about five months working from home when she could, and spending one day a week in the lab.

She is now a fisheries research scientist at the Wisconsin Department of Natural Resources in Madison and is expecting her second child next month. She is not eligible for specific maternity benefits at her new institution, and can only claim the minimum 12 weeks of unpaid leave offered by US law. She plans to use holiday time, sick leave and the 'disability allowance' for which employees become eligible when pregnant to cover most of her salary for about nine of those weeks.

ANNE HODGE

Jenny Briggs occasionally took her daughter with her on fieldwork trips in the early months.

Samantha Joye, an oceanographer at the University of Georgia in Athens, slogged her way through a similarly knotty maternity experience. A tenured professor, she gave birth to her now-6-year-old daughter and worked mainly from home for about 8 months, writing several papers and submitting two grant proposals before her daughter was 6 months old. Because the university did not provide any paid maternity leave, she used sick leave to get partial pay during that time. In 2012, when her twins were born, she did much the same thing. She and her husband, Christof Meile, a biogeochemical modeller at the same university, juggled their parenting around each other's tenure demands.

In the early days, most parents have not yet organized formal child-care services, which presents demands when trying to keep a research programme running. When Joye started venturing back onto campus three months after the birth of her daughter, she brought her child with her when meeting students, attending seminars and joining lab or faculty meetings. Most of her colleagues and university staff were supportive of the presence of her child, she says.

Mara Dierssen, a neuroscientist at the Center for Genomic Regulation in Barcelona, Spain, took about four months of leave after each of her four children were born, but found creative ways to stay in touch with lab members, including walks through the park with her PhD students, with her child in a pushchair.

Hopi Hoekstra, an evolutionary geneticist at Harvard University in Cambridge, Massachusetts, took three months off after her son was born in 2012. But about a month in, she began bringing the baby into the lab, entertaining him in a baby seat and sometimes rocking him to sleep during lab meetings. Her husband, also an evolutionary geneticist at Harvard, took their son into his office as well, where he kept baby gear. In retrospect, she adds, "I wish I had been more off-line. The team did just fine on its own."

Trying to include lab time and face time with graduate students and other lab members during parental leave is difficult enough, but many researchers also have to contend with the demands of fieldwork. Managing that schedule takes some nimble and creative planning, they say (see 'Coping tips'). Jenny Briggs, an ecologist at the US Geological Survey in Denver, Colorado, was working on a project in the Rocky Mountains when her second child was due in 2010. "In my work, biological changes happen season by season," she says. "You can't put annual studies on hold for a year." So she arranged with her boss to funnel the money that she would have been earning had she not taken unpaid leave to a graduate student who could take over the study. She returned to fieldwork just weeks after her daughter was born, sometimes bringing her infant to breastfeed and her mother to assist. Usually, she left the baby at home with a carer and took breaks behind trees or in the field truck to pump breast milk.

TOP NOTCH

Some universities offer exceptional maternity and paternity benefits. Many academic institutions in Scandinavia offer both parents up to

COPING TIPS

How to maximize maternity benefits

Researcher parents have come up with a host of ideas that have helped them to juggle the demands of work and family. Here are a few:

- If possible, hire an assistant or technician to help run lab or field duties during maternity leave.
- Sit down with your department head or chair several months in advance to discuss goals, identify possible support and plan out the return from maternity leave.
- Form a support network of other scientist mums and dads. Facebook groups are one way to build up this type of community. "It's incredibly important," says Jenny Briggs, an ecologist at the US Geological Survey in Denver, Colorado. "Other mothers who do this have been my greatest allies."
- When trying to juggle work and parenting, figure out when you are most productive. "You want to be strategic and realistic about what time of day you're going to work on what," says Natalie Boelman, an Earth scientist at Columbia University in New York. During the early days after her daughter's birth, Boelman found that she needed to sleep during her daughter's morning nap but could write efficiently when her daughter slept in the afternoon.
- As much as possible, share parenting responsibilities 50–50 from day one, says Susana Martinez-Conde, a neuroscientist at the Barrow Neurological Institute in Phoenix, Arizona. She and her husband, Stephen Macknik, also a neuroscientist at Barrow, have shared parenting equally for all three of their children. When necessary, says Martinez-Conde, "we take turns, where one person may be writing a paper or grant and the other takes the kids to the park". A.M.

a year of paid leave, for example, and under UK national policy, institutions must provide statutory maternity pay — a percentage of average weekly earnings — for up to 39 weeks and leave of up to a year. The United States has a few standout examples. At Northwestern University in Evanston, Illinois, many departments offer mothers three months of full pay for childbearing and adoption leave and another three months for child rearing. "I never felt pressure to be in the office during maternity leave," says Yarrow Axford, a geologist who had a son last year. "Of course, I had to keep up with what was happening in my lab and with students conducting independent research. But that is just part of doing the business of science." Northwestern also grants a one-year extension on the tenure clock to mothers after the birth of their child and to both parents following an adoption. Furthermore, it offers three months of paternity leave — so Axford's husband, Christopher Kuzawa, a biological anthropologist at Northwestern, also stayed home for three months after their son's birth. "Those early months are so formative," says Axford. "If more dads could take substantial time away from work to be home with their children in the early months, I suspect there would be more dads out there who are 100% comfortable being left alone with their kids, and moms comfortable leaving them."

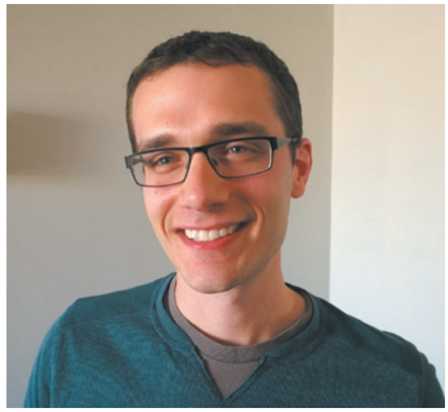
"The scientific community could do a lot about this period when you are pregnant and you cannot be in the lab."
Maria Granberg

Kuzawa notes that the time at home allowed him and Axford to jointly figure out the intricacies of caring for their son. "We established all the patterns together," says Kuzawa. "So it never felt like one of us was the primary caregiver and the other was kind of in the passenger seat." Axford heads out to Greenland this summer for three weeks of fieldwork, but she and her husband feel somewhat less anxious about her impending absence. "We were faced with everything that comes with being a parent — all that middle-of-the-night stuff, weird naps, crying for no reason, won't eat," Kuzawa says. "We had to come up with solutions to all of that. And we did." ■

Amanda Mascarelli *is a freelance writer in Denver, Colorado.*

TURNING POINT
Collin Diedrich

Collin Diedrich has overcome learning disabilities to carve out a promising career researching HIV and tuberculosis (TB) co-infection. He talks about what prompted his move from the United States to the University of Cape Town in South Africa.



What challenges did your disabilities present?
I was diagnosed with reading and learning disabilities in primary school in St Louis, Missouri. But my parents got me private tutors right away, who helped me develop strategies to improve my reading comprehension and organizational and memory skills. Given the stigma and feelings of inadequacy that can come with a learning disability, I have struggled with impostor complex, cycling through phases where I feel completely out of place and inadequate, as if it is only a matter of time until I am 'found out'. Luckily, I am driven and have an amazing support network in my advisers, colleagues, my wife and my family, who have all been helpful and patient when my mind is racing.

What drew you to a career in science?
I didn't really think of science as a career option until I took a biology course at university and realized how much I liked the idea of becoming a researcher. The more I learned about HIV, the more fascinated I was by this virus that can attack your immune system. I read books about it, but what intrigued me most were the first-hand accounts of people with HIV. Before I read them, I couldn't understand why anyone would engage in the risky behaviours that could lead to HIV infection, but I came to understand that the threat of death a decade or more later was not often an immediate concern, especially among those with already-risky lifestyles.

Did you discuss your disabilities with any potential advisers?
When I started my PhD at the University of Pittsburgh, I didn't want to tell people about my learning disorder — I was nervous and intimidated about mentioning it. Then I met JoAnne Flynn, who was head of the molecular virology and microbiology department, and working on simian immunodeficiency virus (SIV) and TB co-infection. Our conversation went so well, I felt comfortable telling her. Without missing a beat, she directed me to the university's learning-disabilities centre. After spending time in her lab, I decided to do my dissertation with her, focusing on the cascade of immunological responses that follows SIV infection. She expected as much from me as from anyone else, and was approachable and helpful — which helped me both scientifically and emotionally.

Why go to South Africa?
I wanted to work on co-infection in human samples, which are in greater supply in South Africa than in the United States. I knew that Robert Wilkinson was working on co-infection at the University of Cape Town, so I secured funding to do my postdoc there.

What did you find hardest about the move?
When you start a postdoc in a new lab, you have to learn where things are and new techniques. I had cultural differences and practical concerns to figure out, and I also had to determine how to get access to the samples necessary to research how HIV alters the granuloma, the inflamed tissue.

How did it change your perspective?
Seeing the effects of these devastating diseases first-hand has been a powerful experience, even though I am doing basic research. I have also learned that just because US scientists do something one way doesn't mean it is the only way. I've replaced my US-centric views with a broader appreciation of research approaches.

What do you plan to do now?
I hope to continue working here throughout my career, at least half-time. I have funding until the end of 2015, and have developed strong collaborations. I would also like to be an advocate for students with learning disabilities and help university admissions officers find ways to look beyond examination scores to include candidates, such as myself, with research aspirations. ■

INTERVIEW BY VIRGINIA GEWIN

THANKS FOR THE MEMORY

It's all in the mind.

BY MILO JAMES FOWLER

Skinner looked like an old vulture with a broken wing. He sat across the kitchen table from Amon with one arm crooked in a scarf bandage — a ruse that, the man insisted, made it easier for him to pass the sentinels with his crisp greenback bribes.

"Boring." Amon blinked, shoulders slumping with disappointment. He tugged the electrodes from his temples, and they retracted on threads spooling from Skinner's open valise. "You promised better."

"They're hard to come by, Sir."

Amon narrowed his hazel eyes. "But you're *easily* replaceable."

Skinner avoided the young Citizen's gaze. "I may have something you'll like." He busied himself with a set of electrodes wired to a separate receptacle inside the black leather bag.

"At additional cost, no doubt."

"Of course." Skinner cringed slightly and bared his yellow teeth in an attempt to lighten the mood. "Got my living to make, same as anybody else. But I can offer you a discount this evening — half off your second purchase."

"Show me what you've got." Amon gestured limply for the smuggler to speed things up. "No more murder and mayhem. I want more *life*."

Electrodes in hand, Skinner reached for the gentleman's forehead.

Amon leaned in. "Is it good?"

"Good, bad — what's the difference?" Skinner forced a chuckle.

Amon grabbed Skinner's gaunt wrist, electrodes dangling in midair. "One word from me, old man, and you will never leave the City. Answer me straight."

"What I mean, Sir, is that the memories themselves are often amoral in nature. A murder, for example. Taken from the mind of the victim's mother, it's the worst possible memory. But from the mind of the killer —"

"I've had quite enough death."

"Understood."

Master Amon had required Skinner's services on multiple occasions over the past few years, ever since the Wall had gone up around the City and President Hoover had mandated that all Citizens remain inside for their own protection. The Hoovervilles outside — as the rabble referred to them — were no place for men of means. But as everyday life behind the Wall stagnated, the wealthy had grown desperate for new sources of entertainment.

"I'm sure you will find this one to be a fine diversion from the norm. If you please."

JACEY

Amon released him. "How'd you come by it?"

"Do you really want to know?" Skinner attached the electrodes to Amon's temples.

Amon held his gaze.

"Very well." Skinner cleared his throat. "Some of the commoners have exhausted their supply of *bad* memories, so to speak. So they have had to resort to releasing *good* ones in light of their … current hardships."

"Winter."

"When it means the difference between buying a warm coat and retaining a lingering memory of Grandma on Christmas morning —" Skinner noted the Citizen's look of disgust. "Not to worry. Considering how well you pay me, there will be no grandmotherly memories for you tonight."

Electrodes in place, Amon leaned back in his chair. "Glad to hear it."

Skinner reached into his bag. "Deep breaths now, Master Amon. You know the drill."

The Citizen shut his eyes and inhaled through his nostrils, waiting for the visions of another life to play through his mind like a motion picture at the theatre — but with full colour and sound.

The sentinel on duty frowned as he halted the Chrysler Six Roadster at the South Wall. Something about the driver didn't look quite right. But as the stoop-shouldered fellow at the wheel reached into his sling, the guard relaxed. When good money changed hands, few questions tended to surface.

"Quite a haul tonight." He pocketed the cash.

The smuggler nodded without saying a word, his fedora pulled low.

Shrugging, the sentinel raised the gate and waved the car through.

Gunning the engine, Amon glanced into the driver's side mirror as he tossed his hat onto the vacant passenger seat. He tugged off the makeshift sling and stretched his arm, reaching back for the bound and gagged man hidden under a throw rug.

"Time to wakey-wakey, Sacagawea." He punched Skinner in the gut and yanked the gag down around his neck, steering one-handed towards the shanties and tents that spread far beyond the reach of the car's headlights.

"Have you lost your mind?" Skinner sputtered.

"I'd be lost without *you*." As the Wall shrank behind them, Amon released a whoop. "I feel so *alive*!"

"What you're feeling now — it's not real. It wasn't your memory —"

"It's mine now. And I'm in love! You hear me?" He howled up at the Moon. "I've got to find her!"

"She won't know you from Madame Butterfly, Sir. And when the G-men find out you've flown the coop —"

Amon howled again.

Skinner sighed. In these harsh economic times, it was best not to give people false hope. Yet that's exactly what this Citizen was riding high on at the moment.

But he would crash back to reality in time. It was inevitable that he would return to his walled city after the throes of true love faded and dropped him into the depths of despair. Amon would suffer a low so devastating he could never hope to recover on his own.

And Skinner would be right there with his valise, when all Amon would demand were *bad* memories again, the variety that never allowed anyone to soar the heavens only to collapse into hell's yawning abyss.

Skinner cleared his throat. "Do you even know her name, Sir?"

Amon scoffed at such a foolish question. He and the woman were soul mates now, after all. "*Florence*," he swooned.

Skinner almost smiled. "Well, that's a start." ■

Milo James Fowler *is a teacher by day, writer by night. When he's not grading papers, he's imagining what the world might be like in dozens of alternate realities. www.milojamesfowler.com*